UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 11-18
Economic Series
May, 2011

Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 916249875

State dependence and heterogeneity in health using a bias corrected fixed effects estimator.*

Jesús M. Carro
Universidad Carlos III de Madrid

Alejandra Traferri
Pontificia Univ. Católica de Chile

Abstract

This paper considers the estimation of a dynamic ordered probit of self-assessed health status with two fixed effects: one in the linear index equation and one in the cut points. The two fixed effects allow us to robustly control for heterogeneity in unobserved health status and in reporting behaviour, even though we can not separate both sources of heterogeneity. The contributions of this paper are twofold. First it contributes to the literature that studies the determinants and dynamics of Self-Assessed Health measures. Second, this paper contributes to the recent literature on bias correction in nonlinear panel data models with fixed effects by applying and studying the finite sample properties of two of the existing proposals to our model. The most direct and easily applicable correction to our model is not the best one, and has important biases in our sample sizes.

JEL classification: C23, C25, I19
Keywords: dynamic ordered probit, fixed effects, self-assessed health, reporting bias, panel data, unobserved heterogeneity, incidental parameters, bias correction.

* This is a revised version of a paper previously circulated under the title "Correcting the bias in the estimation of a dynamic ordered probit with fixed effects of self-assessed health status". We thank Raquel Carrasco, Matilde Machado and seminar participants at Universidad Carlos III, Boston University, and The Lincoln College (Oxford U.) Applied Microeconometrics Conference for useful remarks and suggestions. The first author gratefully acknowledges that this research was supported by a Marie Curie International Outgoing Fellowship within the 7th European Community Framework Programme, by grants number ECO2009-11165 and SEJ2006-05710 from the Spanish Minister of Education, MCINN (Consolider- Ingenio2010) and Consejería de Educación de la Comunidad de Madrid (Excelecon project). The second author would like to thank the Institute for Economic Development at Boston University, in which she conducted part of this research as a visiting scholar. Address for correspondence: Jesús M. Carro. Departmento de Economia, Universidad Carlos III de Madrid, Madrid 28903, Getafe (Madrid-Spain). E-mail addresses: jcarro@eco.uc3m.es; atraferri@uc.cl.

1 Introduction

Self-assessed health (SAH) has been used as a proxy for true overall individual health status in many socioeconomic studies. Moreover, it has been shown to be a good predictor of mortality and of demand for medical care (see, for example, van Doorslaer, Jones, and Koolman, 2004). Motivated by this and by the high observed persistence in health outcomes, Contoyannis, Jones and Rice (2004) study the dynamics and effects of socioeconomic variables on SAH in the British Household Panel Survey. Among other aims, they investigate the relative contribution of state dependence and unobserved heterogeneity in explaining the observed persistence in SAH. State dependence may arise due to structural reasons such as differing abilities to deal with new health shocks depending on previous health status, or willingness to investments in health that changes as health status evolves. For example, people may be less prone to invest in their health after a health shock that lowers their returns to that investment. In any case, as it happens in labor force participation, regardless of the underlying explanations for state dependence, knowing its magnitude is relevant for many health policy debates. This is because the state dependence informs of the long-run implications of a policy affecting health status today.

Given that SAH is a categorical variable Contoyannis, Jones and Rice (2004) use a dynamic ordered probit model, and they take a random effects approach to control for unobserved heterogeneity in the level equation. Halliday (2008) studies the relative contribution of state dependence and unobserved heterogeneity in SAH using a different data set and another random effects approach. Halliday(2008) only includes age as a covariate as the study focuses on the evolution over the life-cycle.

We account for heterogeneity in reporting behavior (cut-point shifts) in addition to heterogeneous unobserved factors that affect health status (index shifts). An example of index shifts is genetic traits. Cut-point shifts occur if individuals use different thresholds to assess their health and report different values of SAH even though they have the same level of true health.[1] Since we can only identify differences up to scale in discrete choice models, we cannot separately identify the two sources of heterogeneity. We can, nonetheless, correctly control for both sources of heterogeneity by including individual effects in the levels and the cut points of the ordered probit. A model with only one individual effect (usually placed in the index equation) allows both sources of heterogeneity too, but so restrictively that it almost always gives incorrect estimates and inferences if both sources are present and relevant.

As with one individual effect, we could take a 'random effects' approach. However, this approach has the drawback of imposing either independence, or a specific and potentially

[1] See Lindeboom and van Doorslaer (2004) for a test that shows evidence of existence of these two different kinds of shifts.

too restrictive functional form on the relation between unobserved heterogeneity and other explanatory variables. It also has the drawback of having to deal with the so-called initial conditions problem. By taking a 'fixed effects' approach, we place no restrictions on the joint distribution of the two individual effects and their correlation with explanatory variables. Moreover, there is no initial conditions problem. Despite these advantages, there have been very few applications of nonlinear panel models with fixed effects in health economics, as noted in Jones' (2007) handbook's chapter.[2] This is due to the known problems in estimating nonlinear panel data models with fixed effects and the panel data sets available. This estimation problem is usually called incidental parameters problem, and it results in large finite sample biases of the MLE when using panels where T is not very large. It is more severe in a model like ours that is dynamic and contains more than one fixed effect.

An important part of the research in microeconometrics has been concerned with finding a solution to this problem by developing bias-adjusted methods. Some examples are Hahn and Newey (2004), Hahn and Kuersteiner (2004), Arellano and Hahn (2006), Carro (2007), Fernandez-Val (2009), and Bester and Hansen (2009).[3] This fast growing literature offers several bias correction methods potentitaly useful to estimate our model. Bester and Hansen (2009) include an application of their so-called HS estimator to a dynamic ordered probit model with two fixed effects. So, the HS is directly applicable to our problem, whereas others require some transformation to adapt them to our model with two fixed effects. However, simulations of other models in the referred papers suggest that HS is not the best one in terms of finite sample performance. They show that for sample sizes with T less than fourteen, the remaining bias when using HS could still be significant, especially for the ordered probit Bester and Hansen (2009) simulate. This result is confirmed in our simulations, which are more specific to the model we want to estimate. Thus, we have to consider another of the proposed methods.

In this paper we derive explicit formulas of the Modified MLE (MMLE) used in Carro (2007) for the dynamic ordered probit model considered here. We evaluate its finite sample performance and compare it with the HS penalty estimator.[4] The MMLE has better finite sample properties and negligible bias in our sample size. This exercise is a main contribution of this paper since, as Arellano and Hahn (2007) point out in their conclusions, more research is needed to know "how well each of the methods recently proposed work for other specific models and data sets of interest in applied econometrics."

[2]Jones and Schurer (2009) is a recent example of using the fixed effects approach to study SAH; however, they use the Conditional MLE of Chamberlain (1980) which does not provide information about the distribution of the fixed effects. This information is needed to calculate marginal effects, the usual parameters of interest in nonlinear models. Another important difference is that Jones and Schurer (2009) do not allow for dynamics.

[3]See Arellano and Hahn (2007) for a good review of this literature, detailed references and a general framework in which the various approaches can be included.

[4]The MMLE comes from modifying the score of the MLE so that the order of the bias in T is reduced.

Also, Greene and Henshen (2008) comment on the lack of studies about the applicability of the recent proposals for bias reduction estimators in binary choice models to ordered choice models.

The rest of the paper proceeds as follows. Section 2 presents our model of SAH, the data we use, and explains the relation of this paper to other recent papers about SAH. Section 3 presents the estimation problem and the method we propose. We also comment on possible solutions from the nonlinear bias correction literature for nonlinear panel data models with fixed effects. We use simulations to evaluate the finite sample performance of different alternatives and to justify selection of MMLE as our estimator. Section 4 presents the estimation results. The estimates of our model and the comparison with random effects estimates show that there are important state dependence effects, and statistically significant effect of income and other socioeconomic variables. Results also show that flexibly accounting for permanent unobserved heterogeneity matters. Section 5 concludes.

2 Model and Data

2.1 Empirical Model of self-assessed health

We consider the following dynamic panel data ordered probit with fixed effects as a reduced-form model of self-assessed health status (SAH):

$$h_{it}^{*} = \alpha_i + \rho_1 \mathbf{1}\left(h_{i,t-1} = 1\right) + \rho_{-1} \mathbf{1}\left(h_{i,t-1} = -1\right) + x_{it}'\beta + \varepsilon_{it};\ i = 1, ..., N,\ t = 0, .., T \quad (1)$$

where x_{it} is a set of exogenous variables that influence SAH, ε_{it} is a time and individual-varying error term which is assumed to be $\varepsilon_{it} \underset{iid}{\sim} N(0,1)$, and h_{it}^{*} is the latent health. The reported SAH (h_{it}), which is what we observe, is determined according to the following thresholds:

$$h_{it} = \begin{cases} -1 & \text{if} \quad h_{it}^{*} < -c_i \\ 0 & \text{if} \quad -c_i < h_{it}^{*} \leq 0 \\ 1 & \text{if} \quad h_{it}^{*} > 0 \end{cases} \quad (2)$$

where $h_{it} = -1$ corresponds to poor health, $h_{it} = 0$ to fair health and $h_{it} = 1$ to good health. α_i and c_i are the model's fixed effects; these account for permanent unobserved heterogeneity, both in unobserved factors affecting health and in reporting behaviour, in an unrestricted way, as explained at the introduction. Note that in addition to the usual scale normalization in discrete choice models (i.e. restricting the variance of ε_{it} to equal one), here we are also normalizing one of the two cut points to be zero. The somewhat more conventional normalization of setting the intercept in the linear index equal to zero is not available to us because the distribution of the intercept, including its mean, is unrestricted in the fixed effects approach. An alternative normalization would be to put the two fixed effects in the two cut points and leave the linear index equation without any intercept.

As this discussion on normalization shows, it is clear that it is not possible to separately identify individual effects affecting that impact only h_{it}^* from those that impact the cut points. Therefore, though we controll for the two mentioned sources of unobserved heterogeneity, we can not separate them. Additionally, having only the fixed effect in the linear index (α_i) would also account for heterogeneity in the cut points, but in a very restrictive way. In particular, by introducing only one individual effect (α_i), we would be assuming that both sources of unobserved heterogeneity must have effects of opposite signs in $\Pr(h_{it} = 1)$ and $\Pr(h_{it} = -1)$; furthermore, we would be restricting how these two effects differ in magnitude for all individuals. We do not have evidence in favor of these assumptions. Furthermore, given the different sources of the unobserved heterogeneity and the potential relations among them and observable variables these assumption are most likely too restrictive, leading to incorrect inference. In contrast with this, by having two fixed effects in (2) we are not imposing any restrictions on the cut-point shifts, nor on the index shift. This constitutes an important difference from previous studies like Contoyannis, Jones and Rice (2004).

In addition to the parameters capturing the effect of heterogeneity, β capture the effect of exogenous variables, and ρ_1 and ρ_{-1} are the parameters that allow state dependence in this model. Determining the relative importance of tate dependence *versus* permanent unobserved heterogeneity as alternative sources of persistence is crucial since they have very different implications. As explained in the introduction, there are several structural reasons for state dependence. However, regardless of the reason, state dependence gives the long-run effect of a policy affecting health status today. This is why it is so useful to know its magnitude.

2.2 Data and x variables

We use the British Household Panel Survey (BHPS), a longitudinal survey of private households in Great Britain. It was designed as an annual survey of each adult (16+) member of a representative sample of more than 5,000 households, with approximately 10,000 individual interviews. The same individuals are re-interviewed in successive waves; if they split off from their original households are re-interviewed along with all adult members of their new households. Similarly, new adult members joining sample households, and children who have reached the age of 16 become eligible for interview. We use sixteen waves of data (years 1991 - 2006), and include individuals who gave a full interview. An unbalanced panel of individuals who were interviewed in at least 8 subsequent waves is used. Our sample consists of 76128 observations from 6,375 individuals.

SAH is defined for waves 1-8 and 10-16 as the response to the question "Compared to people of your own age, would you say your health over the last 12 months on the whole has been: excellent, good, fair, poor, very poor?" In wave 9 the SAH question and categories were reworded. This makes comparison with other waves difficult and wave 9 is not used in our empirical analysis.

The original five SAH categories is collapsed to a three-category variable, creating a new SAH variable that is our dependent variable, with the following codes: poor ($h_{it} = -1$) for individuals who reported either "very poor" or "poor" health; fair ($h_{it} = 0$) for individuals who reported "fair" health; and Good ($h_{it} = 1$) for individuals who reported "good" or "excellent" health.

Main Model. The explanatory variables x that we use in the main model we estimate are: three dummy variables representing marital status (Married, Widowed, Divorced/Separated) with Single as the reference category, size of the household (the number of people living in the same household), number of kids in the household, household income, year dummies (excluding the necessary number to avoid prefect colinearity), and a quadratic function of age. The question about SAH that we use to construct our dependent variable asks respondents to compare health with people their own age. However, SAH becomes worse over time in the raw sample data, perhaps indicating that the age effect over health is not totally discounted by respondents.This can be seen in table 2.[5] This is the reason for including age as an explanatory variable. The income variable is the logarithm of equivalised real income, adjusted using the Retail Price Index and equivalised by the McClement's scale to adjust for household size and composition, and consists on the sum of non-labour and labour income in the reference year.

Variables that are time-constant and specific for individuals, like the level of education or gender, are not included in the set of explanatory variables because they can not be separately identified from permanent unobserved heterogeneity.[6] Fixed effects account for these variables as well as for unobserved characteristics, and we can not separate their effects. Sometimes this is seen as a drawback of the fixed effects approach. However, the random effects approach only separately identifies the effect of these variables because of the unrealistic assumption that unobserved characteristics are independent from them (for example that unobserved healthy life style is independent of education). Even with a correlated random effects approach, if correlation is allowed in a Mundlak (1978) and Chamberlain (1984) style and initial conditions are controlled for following Wooldridge (2005) proposal, it is not possible to separately identify the effect of these time constant variables from the effect of the unobserved factors correlated with them without further assumptions. For instance, Contoyannis, Jones and Rice (2004) follow Wooldridge (2005)

[5] See Contoyannis, Jones and Rice (2004) for further discussion on this.

[6] They are, however, included in the random effects estimation we make for comparison.

proposal, and they comment about this impossibility of separating the effect of variables like education from the effect of the unobservables correlated with them.

Additional Model. In addition to the main model we estimate a model including variables with information on objective health problems. These variables turn in observable part of the unobserved underlying true health, especially persistent health situations. This will help in identifying heterogeneity in reporting behaviour. With this additional model we try to see whether the state dependence that we may find in the main model is still significantly different from zero even after introducing observations of persistent determinants of health. These variables are not clean determinants of SAH and are a mix of several components. Therefore they will induce a decrease in the effect of h_{it-1} even if we correctly capture and isolate all the state dependence effect in the main model. However, if state dependence is still significantly different from zero this will provide further evidence of the robustness and importance of dynamics and state dependence in SAH.

The BHPS contains several questions about health problems and health care demand, but many of them can be induced by a self valuation that might differ from true health as much as SAH, and in an unobserved way. For example the number of visits to the doctor can be determined by a perception of a health problem rather than a true health problem. To avoid this endogeneity bias, we have selected only those questions that we regard as measuring more objective health situations and, therefore, are not affected by personal health assessments. We introduce the following variables:

- Health problems: This is a dummy variable, which takes the value 1 if the individual reports at least one of the following *permanent* health problems or disabilities: arthritis or rheumatism, difficulty in hearing, allergies, asthma, bronchitis, blood pressure, diabetes, migraine or frequent headaches, cancer and stroke, among others.

- Health limits daily activities: This is a dummy variable, which takes the value 1 if the individual answers 'yes' to the following question: does your health in any way limit your daily activities, compared to most people of your age? Examples of daily activities included are: doing the housework, climbing stairs, dressing yourself, walking for at least 10 minutes, etc.

- Health limits ability to work: Similar to previous question.

- Number of days in a hospital as an in-patient in the reference year.

- Finally, we include a dummy variable representing long term sick or disabled, and four other variables for employment status (Self employed, In paid employment, Unemployed, Retired). The category 'Other' (that includes looking after family or home, on maternity leave, on a government training scheme, full-time student/at school, and something else) is left as the reference category.

Table 1: Number of individuals that reports each category of SAH by number of times it is reported.

Number of times	Excellent or good Freq.	%	Fair Freq. (N)	%	Poor or very poor Freq. (N)	%
0	273	4.28	2076	32.56	4380	68.71
1	170	2.67	1114	17.47	898	14.09
2	182	2.85	867	13.60	367	5.76
3	193	3.03	641	10.05	213	3.34
4	233	3.65	481	7.55	137	2.15
5	273	4.28	376	5.90	99	1.55
6	379	5.95	279	4.38	79	1.24
7	456	7.15	204	3.20	46	0.72
8	665	10.43	145	2.27	47	0.74
9	563	8.83	83	1.30	33	0.52
10	533	8.36	61	0.96	32	0.50
11	495	7.76	19	0.30	16	0.25
12	544	8.53	20	0.31	8	0.13
13	672	10.54	5	0.08	9	0.14
14	744	11.67	4	0.06	11	0.17
Total	6375	100.00	6375	100.00	6375	100.00

For example, 273 in the conlumn Freq. of category 'Excellent or good', is the number of individuals that reported 'Excellent or good' 0 times in total over the sample period they are observed.

Descriptive Statistics Tables 1, 2 and 3 contain some descriptive statistics of self-assessed heath reported in our sample. The most frequent category is excellent or good with more than 70% of the answers corresponding to this category. There is high persistence in SAH reported as can be seen in table 3, which shows the transition probabilities. In this table, the largest numbers are on the diagonal for all three values of SAH_{t-1}. Table 2 presents the variation of SAH across different characteristics and health variables. For example, married or single people respond in the excellent or good health category more frequently than widows or divorced people. The three objective health measures in table 2 alter the SAH responses in the expected direction and in greater magnitude than the socioeconomic variables also presented in the table.

2.3 Relation to recent papers studying heterogeneity and state dependence in SAH

2.3.1 Relation to Contoyannis, Jones and Rice (2004)

There is a clear connection between this paper and Contoyannis, Jones and Rice (2004): both papers use the British Household Panel Survey to study the dynamics of SAH. Nev-

Table 2: Proportion (in %) of each category of SAH by several characteristics

Characteristics and their Sample Proportions		SAH categories: Excellent or good	Fair	Poor or very poor
	All	73.19	19.39	7.42
By age group				
40.17	< 40	78.31	16.50	5.19
43.92	40-64	72.92	18.91	8.17
15.91	65+	61.02	28.02	10.96
By sex				
46.84	Male	75.35	18.32	6.34
53.16	Female	71.29	20.34	8.37
By marital status				
63.46	Married	74.00	18.86	7.14
8.92	Divorced	69.63	19.29	11.08
6.32	Widowed	58.84	28.92	12.25
21.3	Single	76.52	18.20	5.28
By household size				
13.30	1	65.57	23.82	10.62
34.32	2	71.67	20.51	7.83
20.20	3	74.33	18.44	7.24
21.63	4	78.30	16.50	5.20
10.55	5+	75.10	17.97	6.93
By kids number				
64.12	0	70.91	20.84	8.25
15.52	1	76.70	17.05	6.25
14.73	2	78.45	16.19	5.36
5.63	3+	75.75	17.74	6.51
Health problems				
58.46	Yes	60.57	27.26	12.16
41.54	No	90.95	8.32	0.74
Health limits daily activities				
13.36	Yes	22.49	39.13	38.38
86.64	No	81.01	16.35	2.64
Health limits work				
16.43	Yes	29.85	38.29	31.86
83.57	No	81.71	15.68	2.61

Table 3: Sample transition probabilities from SAH in t-1 to SAH in t

		SAH in t			
		Excellent or good	Fair	Poor or very poor	Total
SAH	Excellent	85.91	11.84	2.25	100
in	Fair	43.22	45.18	11.59	100
$t-1$	Poor or very poor	17.66	31.60	50.74	100
	Proportion	72.80	19.67	7.53	100

ertheless, there are several aspects considered in Contoyannis, Jones and Rice (2004) that are not studied here. In particular, that paper contains a more detailed data description, and a further discussion of the estimated model; it also address other issues, like sample attrition, that are not considered here.[7] However, our paper complements and adds to Contoyannis, Jones and Rice (2004) in various ways:

(i) We use more periods from the BHPS than they do. They only use the first eight waves because the ninth contains a different question about and categorization of SAH. While we drop the 9th wave too, we incorporate the waves after the 9th in our estimation. Since the model specified includes only one lag of h_{it}, we have all the variables we need for the 11th to 16th waves. For the 10th wave we have all the variables but h_{it-1} as it is the case for the first wave. We treat the 10th wave like an initial observation and condition it out in our likelihood leaving the probability of that observation totally unrestricted. Contoyannis, Jones and Rice (2004) can not do this because of their way of solving the initial conditions problem and use of random effects.

(ii) In our model we have two individual specific effects: one in the linear index and one in the cut points. Lindeboom and van Doorslaer (2004) tested for and found clear evidence of different reporting behavior (cut-point shifting) for gender and age. Given that Contoyannis, Jones and Rice (2004) impose homogeneous cut points, they estimate different models by gender to allow for that differing reporting behavior, but they do not allow unrestricted different behavior by age. Although we can not separately identify both sources of unobserved heterogeneity, our approach is robust to heterogeneous cut points freely correlated with any determinant of SAH.

(iii) We use fixed effects instead of a random effects approach. The advantages of this are

[7] An unbalanced panel (with random attrition) in a dynamic panel model does not pose any complications to a fixed effect estimator (as opposed to a random effects estimator), as long as it does not imply many individuals with a very small number of periods; and in our sample all observations have at least 8 periods. However, the assumption of attrition at random seems unrealistic. Contoyannis, Jones and Rice (2004) made a test and found evidence of non-random attrition, but they also found that the bias this may be causing to the estimates is negligible. Given this result using the same data set as us, and since this problem would take us too far from the main theme of this paper, we do not consider it here.

that no arbitrary restriction is imposed on the correlation between permanent unobserved heterogeneity and the observable variables, and there is no initial conditions problem.

(iv) As an additional complement, our study includes some objective health measures, so we can see how much is explained by socioeconomic variables and by state dependence even after these measures are included.

Given the aspects not covered in this paper, and in order to make an assessment of the contributions of this paper with respect to the previous literature we also estimate our models using the same kind of specification and estimation method as Contoyannis, Jones and Rice (2004). Thus that we also estimate (2) using a correlated random effects specification with only an individual effect in the linear index equation (the α_i parameter in (1)), but with homogeneous cut points. Therefore, in this correlated random effects specification:

$$h_{it} = \begin{cases} -1 & \text{if} \quad h_{it}^* < c_1 \\ 0 & \text{if} \quad c_1 < h_{it}^* \leq c_2 \\ 1 & \text{if} \quad h_{it}^* > c_2 \end{cases} \tag{3}$$

where c_1 and c_2 are (homogenous) parameters to be estimated, h_{it}^* is defined in (1), and α_i in (1) is assumed to be:

$$\alpha_i = \gamma_0 + \gamma_1' h_{i1} + \gamma_2' \overline{x}_i + u_i \tag{4}$$

where $\overline{x}_i$ is the average over the sample period of the exogenous variables, and $u_i \underset{iid}{\sim} N(0, \sigma_u^2)$ independently of everything else. h_{i1} is in (4) to deal with the initial condition problem following Wooldridge (2005).

2.3.2 Relation to Halliday (2008)

Halliday (2008) studies state dependence and heterogeneity in SAH using data from the Panel Study of Income Dynamics. Since his focus is on the evolution of health over the life-cycle, he only considers age as explanatory variable. No other socio-economic variable is included. Another difference with our study is that he further reduces health status to two categories estimating a logit instead of an ordered probit with three categories. With respect to heterogeneity, on the one hand Halliday (2008) is more flexible because his analysis allows for heterogeneous parameters both in the intercept and in the slope of the dynamic model. On the other hand the random effects approach he adopts has no incidental parameters problem but restricts the distribution of the heterogeneity and suffers from the initial conditions problem. Nonetheless, Halliday (2008) uses a discrete finite mixture, which is potentially more flexible and less parametric in its treatment of heterogeneity and the initial conditions than the distribution assumed in Contoyannis, Jones and Rice (2004). This greater flexibility comes from the possibility of having many points of support which should provide an approximation to a variety of distributions like asymmetric distributions, or distributions with several modes.

The limitation of this approach is computational. This limitation leads Halliday (2008) to consider no more than four points of support, even though more points of support might be needed.[8] Four is certainly more than the two points of support assumed in other applications, but it may be not enough to provide a good approximation to a bivariate distribution. In this paper we have two individual specific effects potentially correlated with each other. Such a bivariate joint distribution may be difficult to approximate with only four points of support. In contrast to these limitations, the fixed effects approach we follow here is non-parametric in the distribution of the heterogeneity, requires no special treatment of the initial conditions, and does not have the same computational limitations as estimating discrete finite distributions.

Finally, Halliday (2008) finds evidence of a great amount of heterogeneity in health, which is the most important motivation for following the approach we propose here.

2.3.3 Relation to Jones and Schurer (2009)

Another recent paper dealing with self-assessed health measures is Jones and Schurer (2009). They use the German Socio-Economic Panel and focus on the effect of income on health, conducting a detailed analysis on the shape of this effect. But, it does not consider dynamics in SAH. However, it does address the potential importance of unobserved heterogeneity. They control for heterogeneity in both unobserved health status and reporting behaviour by estimating a logit model with fixed effects for each of the $J - 1$ threshold values into which SAH can be dichotomized. However, they estimate it using the Conditional MLE of Chamberlain (1980) because the standard MLE estimation of ordered-choice models with fixed effects suffer from a severe incidental parameters problem and there was no other solution from the panel data literature ready to be applied to these models. Implementing a solution to this problem in the estimation of ordered-choices model with fixed effects is one of the main contributions of our paper. This allow us to estimate marginal effects that properly account for the distribution of unobserved heterogeneity and, especially, for its correlation with observable variables. The Conditional MLE conditions out the fixed effects and, therefore, no information about them is recovered. This means that when calculating marginal effects they have to substitute the fixed effect for a value that may not be representative of the population and, in any case, it ignores the correlation between the observables and the heterogeneity.

Jones and Schurer (2009) also estimate a random effects model that assumed independence of the heterogeneity. Comparing this with the fixed effects estimates they find

[8] See section 5.1.2 in Halliday (2008).

that the underlying assumptions of the statistical model matter for assessing the link between income and health. This finding provides additional support in favor of estimating a model that makes no assumptions about the distribution of the heterogeneity.

3 Estimation Method

3.1 Estimation problem and possible solutions

From (1), (2) and the normality assumption about ε_{it}, we have that

$$\Pr(h_{it} = -1|x_{it}, h_{it-1}, c_i, \alpha_i) = 1 - \Phi\left(c_i + \mu_{it}\right) \tag{5}$$

$$\Pr(h_{it} = 0|x_{it}, h_{it-1}, c_i, \alpha_i) = \Phi\left(c_i + \mu_{it}\right) - \Phi\left(\mu_{it}\right)$$

$$\Pr(h_{it} = 1|x_{it}, h_{it-1}, c_i, \alpha_i) = 1 - \Pr(h_{it} = -1|.) - \Pr(h_{it} = 0|.) = \Phi\left(\mu_{it}\right) \tag{6}$$

where

$$\mu_{it} = \alpha_i + \rho_1 \mathbf{1}\left(h_{i,t-1} = 1\right) + \rho_{-1}\mathbf{1}\left(h_{i,t-1} = -1\right) + x_{it}'\beta \tag{7}$$

Conditioning on the first observation h_{i0}, the log-likelihood is:

$$\begin{aligned} l(\rho_1, \rho_{-1}, \beta, \alpha, \mathbf{c}) = \sum_{i=1}^{N}\sum_{t=1}^{T} \{ & \mathbf{1}\left\{h_{it} = -1\right\} \log\left[1 - \Phi\left(c_i + \mu_{it}\right)\right] + \\ & \mathbf{1}\left\{h_{it} = 0\right\} \log\left[\Phi\left(c_i + \mu_{it}\right) - \Phi\left(\mu_{it}\right)\right] + \mathbf{1}\left\{y_{it} = 1\right\} \log\left[\Phi\left(\mu_{it}\right)\right]\}, \end{aligned} \tag{8}$$

Using standard MLE to estimate models like (2) is known to be biased, since we do not have a large number of periods. The MLE is inconsistent when T does not go to infinity because the fixed effects act as incidental parameters. Furthermore, existing Monte Carlo experiments with dynamic nonlinear models show that the MLE has large bias. In fact, simulations of a dynamic ordered probit in Bester and Hansen (2009) and simulations in the following sections show that the bias is non-negligible even with T as large as 20. As mentioned in the introduction, several recently developed bias-correction methods could overcome this problem. Arellano and Hahn (2007) summarize the different approaches.

The methods can be grouped into three approaches based on the object that is corrected. The first aproach is to construct an analytical or numerical bias correction of a fixed effect estimator. Fernandez-Val (2009), among others, takes this approach and applies his analytical bias correction to dynamic binary choice models. The second approach is to correct the bias in moment equations. An example of this is Carro (2007) that uses an estimator of this type to correct the bias in dynamic binary choice models. The third group are those that correct the objective function. Arellano and Hahn (2006) and Bester and Hansen (2009) take this approach, with the latter including an application to a dynamic ordered probit model. The HS-penalty estimator studied in Bester and

Hansen (2009) is the first option we consider because our model is also a dynamic ordered probit, and because alternative approaches require transformations or derivations. This estimator also has the advantage of being easier to compute than the Modified MLE in Carro (2007) and the Bias Correction in Fernandez-Val (2009) because the HS does not require the calculation of expectations and the other two do. This advantage is more relevant in our case, because it has two fixed effects.

Arellano and Hahn (2007) show how the different approaches are related. Asymptotically, all the approaches always reduce the order of the bias of the MLE from the standard $O(T^{-1})$ to $O(T^{-2})$ for the general classes of models they were developed. However there may be differences when they are applied to specific cases . The following very simple example used in Carro (2007), Arellano and Hahn (2007), and Bester and Hansen (2009) illustrates this point. Consider the model where $y_{it} \underset{iid}{\sim} N(\eta_i, \sigma_0^2)$. The ML estimator of σ_0^2 is $\widehat{\sigma}^2_{MLE} = \frac{1}{NT}\sum_i \sum_t (y_{it} - \widehat{\eta}_i)^2$. It is well known that $\widehat{\sigma}^2_{MLE}$ is not a consistent estimator of σ_0^2 when $N \to \infty$ with fixed T, since it converges to $\frac{T-1}{T}\sigma_0^2$. In this case the whole problem is very easy to fix. $\frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \widehat{\eta}_i)^2$ is the fixed T consistent estimator of σ_0^2. The MMLE from Carro (2007) produces this very same estimator, correcting not only the $O(T^{-1})$ term of the bias, but all the asymptotic bias in this special example. The HS removes the $O(T^{-1})$ term of the bias, but it does not attain the fixed-T consistent estimator. The one-step bias correction to the ML estimator from Fernandez-Val (2009) does not produce a fixed-T consistent estimator either, but its iterated form does. Thus, differences may appear between the different approaches when applied to specific models.

On the other hand, the incidental parameters problem can be seen as a finite sample bias problem in panel data context. The problem is not important when T is large relative to N. However, since our panel does not have a large number of periods it is reasonable to wonder whether the excellent asymptotic properties of the MLE when T goes to infinity (sufficiently fast) are a good approximation to our finite sample. Simulations show that we would need panels with many more time periods than are usually found in practice. The relevant implication is that we have to examine the finite sample performance of the estimators for our model and sample sizes. In the methods considered here this is done through Monte Carlo experiments. Bester and Hansen (2009) do not compare the finite sample properties of the method they use with others for the ordered probit case because many of the other methods require some derivation to get the specific correction for this case. However, they make such a comparison using binary choice (probit and logit) models. Also, Carro (2007) and Fernandez-Val (2009) conduct Monte Carlo experiments for logit and probit models with different sample sizes (both in T and N), allowing us to compare a wide range of methods for those models. From these comparisons we can conclude that the HS penalty approach is not the best one and for sample sizes with T smaller than 13 the remaining bias can still be significant. Given this result, we consider other of the proposed methods to estimate our ordered probit and evaluate its finite sample properties.

Interesting candidates are the corrections discussed by Fernandez-Val (2009) and Carro (2007) since they are equally superior to other alternatives in finite sample performance in the relevant existing comparisons. In the next subsections we derive explicit formulas of the modified MLE used in Carro (2007) for the model considered here, evaluate its finite sample performance, and compare it with the HS penalty estimator.

3.2 MMLE for a dynamic ordered probit with two fixed effects

The model we want to estimate is defined in (1) and (2), and its log-likelihood is (8). Let $\gamma = (\beta, \rho_1, \rho_{-1})$ and $\eta_i = (\alpha_i, c_i)$. Partial derivatives are denoted by the letter d, so the first order conditions are $\mathbf{d}_{\eta i}(\gamma, \eta_i) \equiv \frac{\partial l_i(\gamma,\eta_i)}{\partial \eta_i}$ and $\mathbf{d}_{\gamma i}(\gamma, \eta_i) \equiv \frac{\partial l_i(\gamma,\eta_i)}{\partial \gamma}$. Bold letters represent vectors.

The MLE of η_i for given γ, $\eta_i(\gamma)$, solves $\mathbf{d}_{\eta i}(\gamma, \eta_i) = 0$. The MLE of γ is obtained by maximizing the concentrated log-likelihood ($\sum_{i=1}^{N} l_i(\gamma, \eta_i(\gamma))$), i.e. by solving the following first order condition:

$$\frac{1}{T\,N} \sum_{i=1}^{N} \mathbf{d}_{\gamma i}(\gamma, \eta_i(\gamma)) = 0 \tag{9}$$

where $\mathbf{d}_{\gamma i}(\gamma, \eta_i(\gamma)) = \left.\frac{\partial l_i(\gamma,\eta_i)}{\partial \gamma}\right|_{\eta_i=\eta_i(\gamma)}$.

To reduce the bias of the estimation, we follow Carro (2007) in modifying the score of the concentrated log-likelihood by adding a term that removes the first order term of the asymptotic bias in T. By doing so, we get that the MMLE of the γ parameters of model (2) is the value that solves the following score equation:

$$\begin{aligned}
\mathbf{d}_{\gamma M i}(\gamma) = \mathbf{d}_{\gamma i}(\gamma, \eta_i(\gamma)) - \frac{1}{2}\frac{1}{d_{\alpha\alpha i}d_{cci} - {d_{\alpha ci}}^2}\Bigg[d_{\alpha\alpha i}\left(\mathbf{d}_{\gamma cci} + d_{\alpha cci}\frac{\partial \widehat{\alpha}_i}{\partial \gamma} + d_{ccci}\frac{\partial \hat{c}_i}{\partial \gamma}\right) \\
+ d_{cci}\left(\mathbf{d}_{\gamma\alpha\alpha i} + d_{\alpha\alpha\alpha i}\frac{\partial \widehat{\alpha}_i}{\partial \gamma} + d_{\alpha\alpha ci}\frac{\partial \hat{c}_i}{\partial \gamma}\right) - 2d_{\alpha ci}\left(\mathbf{d}_{\gamma\alpha ci} + d_{\alpha\alpha ci}\frac{\partial \widehat{\alpha}_i}{\partial \gamma} + d_{\alpha cci}\frac{\partial \hat{c}_i}{\partial \gamma}\right)\Bigg] \\
- \frac{\partial}{\partial \alpha_i}\left.\left(\frac{E(\mathbf{d}_{\gamma ci})E(d_{\alpha ci}) - E(d_{cci})E(\mathbf{d}_{\gamma\alpha i})}{E(d_{\alpha\alpha i})E(d_{cci}) - [E(d_{\alpha ci})]^2}\right)\right|_{\eta_i=\eta_i(\gamma)} \\
- \frac{\partial}{\partial c_i}\left.\left(\frac{E(\mathbf{d}_{\gamma\alpha i})E(d_{\alpha ci}) - E(d_{\alpha\alpha i})E(\mathbf{d}_{\gamma ci})}{E(d_{\alpha\alpha i})E(d_{cci}) - [E(d_{\alpha ci})]^2}\right)\right|_{\eta_i=\eta_i(\gamma)} = 0
\end{aligned} \tag{10}$$

where $\mathbf{d}_{\gamma i}(\gamma, \eta_i(\gamma))$ is the standard first order condition from the concentrated log-likelihood, as in (9). $\mathbf{d}_{\gamma ci} = \frac{\partial^2 l_i}{\partial \gamma \partial c_i}$, $d_{\alpha\alpha i} = \frac{\partial^2 l_i}{\partial \alpha_i^2}$, $\mathbf{d}_{\gamma\alpha ci} = \frac{\partial^3 l_i}{\partial \gamma \partial c_i \partial \alpha_i}$, and so on. From the first order conditions of α_i and c_i we obtain $\widehat{\alpha}_i(\gamma)$ and $\widehat{c}_i(\gamma)$, as it is done in order to concentrate the log-likelihood. All expectations are conditional on the same set of information as the likelihood. These expectations can be computed by conditioning recursively, like we do to write the conditional likelihood. The parametric model (equations (1), (2) and the assumption about ε_{it}) from which we write the likelihood also gives the parametric form of the expectations we need to calculate.[9]

We show in Appendix A how this modification on the score of the concentrated log-likelihood in (10) is a first order adjustment on the asymptotic bias of the ML score, so the first order condition is more nearly unbiased and the order of the bias of the estimator is reduced from $O(T^{-1})$ to $O(T^{-2})$. Furthermore, the bias is corrected without changing the asymptotic variance of the MLE.

3.3 Simulations

3.3.1 First DGP: Performance for different T

We simulate the model in equations (1), and (2) with the following value of the parameters and Data Generating Process (DGP): $\beta = 1$, $\rho_1 = 0.5$, and $\rho_{-1} = -0.5$. The error follows a normal distribution: $\varepsilon_{it} \sim N(0, 1)$. The fixed effects are constructed as follows:

$$\alpha_i = \frac{1}{2}\sum_{t=1}^{4} x_{it} + u_i, \quad \text{where } u_i \sim N(x_{i0}, 1) \tag{11}$$

$$c_i = |z_i|, \quad \text{where } z_i \sim N(x_{i0}, 1). \tag{12}$$

so that they are correlated with the explanatory variables. This correlation of the unobserved heterogeneity with the covariates makes the problem more severe than in the independency case. We study the performance of estimators under this condition as we consider it to be more realistic.[10] x_{it} follows a Gaussian AR(1) with autoregressive parameter equal to 0.5. Initial conditions are $x_{i0} \sim N(0, 1)$ and $h^*_{i0} = \alpha_i + \beta_0\, x_{i0} + \varepsilon_{i0}$. We perform 1000 replications, with a population of $N = 250$ individuals. For each simulation we estimate the MLE, the MMLE given by equation (10) and the HS estimator defined in Bester and Hansen (2009). That is, the HS estimator is the value of the parameters that maximize the following penalized objective function:

$$\sum_{i=1}^{N} lk_i\left(\beta, \rho_1, \rho_{-1}, \alpha_i, c_i\right) - \sum_{i=1}^{N}\frac{1}{2} trace\left(\widehat{I}_{\alpha c_i}^{-1}\widehat{V}_{\alpha c_i}\right) - \frac{k}{2} \tag{13}$$

where lk_i is the log likelihood of i, $\widehat{I}_{\alpha c_i}$ is the sample information matrix for $e_i = (\alpha_i, c_i)'$, $\widehat{V}_{\alpha c_i}$ is a HAC estimator of $Var\left(\frac{1}{\sqrt{T}}\frac{\partial l_i}{\partial e_i}\right)$, and $k = \dim(e_i)$. This penalty term is easier to calculate than the modification of the score in (10) because the penalty does not involve any expectation.

[9] Appendix B gives some indications about computing the MMLE.

[10] In the simulations of an ordered probit in Bester and Hansen (2009) the fixed effects are independent of the covariates. We have simulated and compared MMLE and HS in this case too. As said, the bias is smaller for all T, but the conclusions from the comparison between MMLE and HS are the same as in the dependency case. Since the latter is more relevant in practice we do not report the independency case.

Table 4: Monte Carlo Results. Dynamic Ordered Probit parameters

Parameter	β		ρ_1		ρ_{-1}	
True value	1		0.5		-0.5	
Estimator	Mean Bias	RMSE	Mean Bias.	RMSE	Mean Bias	RMSE
			$T = 4$			
MLE	0.816	0.828	-0.474	0.516	0.551	0.586
HS	0.796	0.809	-0.392	0.443	0.467	0.509
MMLE	0.172	0.182	-0.254	0.282	0.280	0.305
			$T = 8$			
MLE	0.335	0.341	-0.188	0.216	0.189	0.216
HS	0.247	0.254	-0.115	0.153	0.119	0.154
MMLE	0.073	0.086	-0.062	0.108	0.067	0.109
			$T = 10$			
MLE	0.257	0.263	-0.145	0.171	0.154	0.179
HS	0.170	0.178	-0.083	0.119	0.093	0.127
MMLE	0.052	0.067	-0.036	0.086	0.050	0.093
			$T = 12$			
MLE	0.210	0.215	-0.127	0.152	0.127	0.151
HS	0.127	0.134	-0.072	0.106	0.074	0.106
MMLE	0.040	0.054	-0.030	0.079	0.036	0.081
			$T = 16$			
MLE	0.154	0.159	-0.093	0.118	0.096	0.119
HS	0.081	0.088	-0.048	0.083	0.054	0.085
MMLE	0.026	0.041	-0.017	0.068	0.022	0.069
			$T = 20$			
MLE	0.122	0.127	-0.072	0.095	0.078	0.101
HS	0.058	0.065	-0.034	0.067	0.042	0.074
MMLE	0.019	0.034	-0.009	0.058	0.016	0.062

Note: See a detailed description of the model simulated and other characteristics of the DGP in subsection 3.3.

Results from this experiment for different T are reported in Table 4, which shows the mean bias and the Root Mean Squared Error (RMSE). We find that for all T, the MMLE performs much better than the other two estimators. Comparing it with the HS, the differences are greater for $T = 4$ and $T = 8$, where the HS is closer to the MLE than to the MMLE. When using the MMLE the bias is smaller than 10% of the true values with $T = 10$ for all but for one of the ρ parameters. With $T = 12$ the bias when using the MMLE is already negligible whereas the HS contain biases and RMSE larger than the MMLE with $T = 10$. Even with $T = 16$ the HS exhibit mean biases greater than the MMLE with $T = 10$. It is not until $T = 20$ that the HS has small biases and RMSE. So HS needs more periods (at least more than 16) to have small finite sample biases. Given this and the fact that the sample sizes we have in our empirical analysis are smaller than $T = 14$, we use MMLE.

The reasons of this better performance of the MMLE is the use of the specific structure of the model we want to estimate, which translates into the expectations in the modification term. The likelihood includes the fact that we know the distribution of one of the explanatory variables: the lag of the dependent variable. This distribution is, of course, that of the dependent variable in the previous period. Therefore, we write the likelihood recursively for each period (conditional on the previous period) up to the likelihood of the initial condition. This is used in the modification so it includes expectations, using the known distribution of h_{it-1} conditional on h_{it-2}. The HS is generally written so that it does not make any intensive use of a specific likelihood and it does not include such expectations. Therefore HS does not exploit all the information that our specification provides and it requires more periods to attain the same performance as the MMLE. This confirms the idea expressed in Bester and Hansen (2009) that the simplicity of the HS (due to not having to calculate expectations) may not be free and could lead to a worse performance than other approaches.

3.3.2 Quality of inference

We consider the quality of inference on finite samples based on these estimators. Table 5 presents the coverage of 95% confidence intervals and the estimated asymptotic standard errors divided by the standard deviation. The latter is very close to 1 in all cases for the MMLE and in most cases for the other estimators, which indicates that the variance is estimated well and the problem is the bias. This corresponds with the fact that we are correcting a bias without altering the asymptotic variance. With respect to inference, the coverage of the confidence intervals is extremely poor for the MLE, especially for β. Even with $T = 20$, the coverage for β is smaller than 3%. The HS estimator improves inference with respect to the MLE, but it is still too far from the theoretical coverage of 95%, being the coverage for β specially bad even with $T = 20$. As it happens with the bias and RMSE criteria, the MMLE is clearly the best estimator of these three for doing inference, for all periods and parameters.

3.3.3 Performance for different degrees of persistence

To check whether results are maintained under different state dependence scenarios, we present simulations for different values of ρ_1 and ρ_{-1}, with $T = 10$ in Table 6. The DGP is the same as that of Table 4 except for the values of ρ_1 and ρ_{-1}. Here the state dependence changes from very negative to very positive, including the case with no state dependence. In terms of bias and RMSE, we find that the MMLE performs better than the other methods for all cases. In principle, having a more negative state dependence may improve all the estimators since it induces higher variance in y_{it}. This is the case for the estimation of β, where the three estimation methods improve, but it is not the case

Table 5: Monte Carlo Results. Inference over Dynamic Order Probit parameters: Conference intervals coverage and estimation of the estandard error.

Parameter	β		ρ_1		ρ_{-1}	
True value	1		0.5		-0.5	
Estimator	% Coverage C.I. 95%	SE/SD	% Coverage C.I. 95%	SE/SD	% Coverage C.I. 95%	SE/SD
			$T = 8$			
MLE	0%	0.85	47%	0.87	48%	0.90
HS	0%	0.86	74%	0.91	73%	0.94
MMLE	64%	1.02	87%	0.93	85%	0.96
			$T = 10$			
MLE	0%	0.81	54%	0.91	53%	0.91
HS	3.5%	0.83	82%	0.96	78%	0.95
MMLE	74%	0.94	90%	0.96	89%	0.96
			$T = 12$			
MLE	0%	0.89	58%	0.91	62%	0.93
HS	8.8%	0.92	85%	0.96	83%	0.98
MMLE	81%	1.00	92%	0.95	92%	0.97
			$T = 16$			
MLE	0%	0.92	69%	0.91	68%	0.94
HS	29%	0.95	88%	0.96	88%	0.99
MMLE	88%	1.00	93%	0.94	93%	0.96
			$T = 20$			
MLE	2%	0.90	77%	0.96	73%	0.94
HS	48%	0.93	91%	1	88%	0.98
MMLE	90%	0.97	95%	0.98	93%	0.95

Note: This is for the simulation experiment in Table 4. We have used the inverse of the hessian as estimator of variance.

for the estimation of ρ_1 and ρ_{-1}, where the MMLE improves but the MLE and HS get worse.

3.3.4 Simulations based on real data

Finally, we perform a simulation based on the real data used in this paper. This will provide further evidence about finite sample performance of the MMLE and will give more robustness to our estimator choice. The DGP takes the estimates obtained by MMLE and reported in Table 8 as the true model. It takes the real data for all the individuals used in that estimation and all the significant x variables except the time dummies. This means that in this DGP x_{it} is a vector containing observations of the following variables: age, squared age, household size, number of kids, and income. The true values of the parameters are: $\rho_1 = 0.4875$, $\rho_{-1} = -0.4375$, $\beta' = (0.0205, -0.0005, -0.0388, 0.0472, 0.0396)$. $N = 1739$, T is the same as in our data (i.e. between 8 and 14 periods), and $\varepsilon_{it} \sim N(0, 1)$.

Table 6: Monte Carlo Results. Dynamic Ordered Probit parameters with different degrees of state dependence

Parameter	β		ρ_1		ρ_{-1}	
Estimator	Mean Bias	RMSE	Mean Bias.	RMSE	Mean Bias	RMSE
True value	1		-1		1	
MLE	0.204	0.212	-0.264	0.284	0.244	0.265
HS	0.105	0.116	-0.094	0.136	0.087	0.130
MMLE	0.012	0.044	-0.008	0.089	0.003	0.086
True value	1		-0.5		0.5	
MLE	0.212	0.218	-0.214	0.235	0.206	0.227
HS	0.116	0.126	-0.079	0.119	0.078	0.119
MMLE	0.026	0.048	-0.018	0.083	0.018	0.083
True value	1		0		0	
MLE	0.227	0.233	-0.180	0.201	0.180	0.201
HS	0.136	0.144	-0.079	0.116	0.084	0.119
MMLE	0.037	0.055	-0.028	0.082	0.032	0.084
True value	1		0.5		-0.5	
MLE	0.257	0.263	-0.145	0.171	0.154	0.179
HS	0.170	0.178	-0.083	0.119	0.093	0.127
MMLE	0.052	0.067	-0.036	0.086	0.050	0.093
True value	1		1		-1	
MLE	0.297	0.303	-0.105	0.144	0.111	0.148
HS	0.215	0.222	-0.086	0.126	0.091	0.129
MMLE	0.065	0.078	-0.057	0.100	0.069	0.107

Note: 1000 Monte Carlo simulations of the Ordered Probit model in equations (1) and (2), following the same DGP as in Table 4 (described at the beginning of section 3.3), but changing the value of the state dependence parameters from negative to positive, including the case with no state dependence. $T = 10$.

α_i and c_i are the estimates of these parameters by MMLE. The distributions of these two parameters can be seen in graph 1. The distribution of α_i is not normal and is correlated with c_i (correlation coefficient between α_i and c_i is -0.33). Thus, the distribution of unobserved heterogeneity is not an arbitrary and statistically convenient distribution, but an empirically founded distribution that captures both real correlations with the covariates and correlations between fixed effects. These correlations and distributions of α_i and c_i are richer than those in the previous simulation experiments. Furthermore, this is the relevant DGP to compare the proposed strategy for dealing with unobserved heterogeneity with the random effects approach previously used in the literature. Making this comparison with an arbitrarily chosen DGP may imply a too favorable assumption to the random effects, as in our first DGP, or a too arbitrarily unfavorable one. However, this case is the relevant case for our empirical analysis.

For the reasons discussed, we evaluate the finite sample performance of the random

Table 7: Monte Carlo Results. DGP based on the real data used in the empirical analysis.

	ρ_{-1}	ρ_1	Age	Age^2	House-hold size	Number of Kids	Household Income
True value	-0.4375	0.4875	0.0205	-0.0005	-0.0388	0.0472	0.0396
			Mean Bias				
CRE	0.0945	0.0459	0.0002	0.00006	-0.0080	0.0095	-0.0003
MLE	0.2039	-0.1239	0.0061	-0.00016	-0.0078	0.0121	0.0063
HS	0.1288	-0.0474	0.0044	-0.00010	-0.0049	0.0077	0.0033
MMLE	0.0437	0.0090	0.0029	-0.00006	-0.0030	0.0046	0.0016
			Root Mean Squared Error				
CRE	0.1041	0.0603	0.0265	0.00021	0.0406	0.0512	0.0348
MLE	0.2066	0.1272	0.0113	0.00018	0.0304	0.0354	0.0257
HS	0.1326	0.0546	0.0098	0.00013	0.0271	0.0310	0.0234
MMLE	0.0576	0.0289	0.0086	0.00010	0.0261	0.0292	0.0222

Note: 1000 Monte Carlo simulations. DGP described at the begging of subsection 3.3.4.

effects approach (CRE) described at the end of section 2.3.1, in addition to the MLE, HS, and MMLE. To make the comparison as close as possible with the estimators used in practice, we include the following constant variables as covariates when estimating by random effects: gender, race, and education indicators. These are implicitly included in the DGP through the estimated α_i and c_i, since in the fixed effects these variables can not be separately identified from the fixed effects.

The results of this simulation are presented in Table 7. The MMLE is clearly the best of all estimators in terms of RMSE. More specifically, the bias and RMSE for the CRE are twice the bias and RMSE of the MMLE for some parameters like ρ_1 and Household Size. As in the previous simulations experiments with similar number of periods, the MMLE exibit small biases.

4 Estimation Results

4.1 Main Model

Table 8 presents the coefficient estimates for the main model based on three different estimators. This includes different specifications of the heterogeneity. The first estimated model (column I) is a pooled version of the model in (1) and (2), without individual specific effects. The second estimated model (column II) is the correlated random effects model described in equations (3) and (4). It is similar to models estimated in Contoyannis, Jones and Rice (2004). It has homogenous cut-points and uses a random effects approach to control for the individual specific intercept in the linear index. The last specification (column III) is described in previous subsections; it is the model in (1) and (2) treating

Table 8: Estimates, Main model.

Variables	I Pooled	II Correlated Random Effects	III MMLE
Health in t-1: Good	0.6527*** (0.0185)	0.5028*** (0.0234)	0.4875*** (0.0186)
Health in t-1: Poor	-0.4417*** (0.0233)	-0.3259*** (0.0343)	-0.4375*** (0.0242)
Age	0.0011 (0.0032)	0.0200 (0.0210)	0.0205 (0.0222)
Age square	-0.0000 (0.0000)	-0.0007*** (0.0001)	-0.0005*** (0.0001)
Married	0.0344 (0.0286)	0.1722 (0.0752)	0.0749 (0.0606)
Separated/Divorced	-0.0580 (0.0358)	0.0475 (0.1028)	0.0375 (0.0729)
Widowed	-0.0243 (0.0408)	0.3668** (0.1329)	0.0542 (0.0918)
Household size	-0.0782*** (0.0138)	-0.0112 (0.0189)	-0.0388** (0.0177)
Number of Kids	0.0647*** (0.0155)	0.0423 (0.0189)	0.0472** (0.0188)
Household Income	0.0816*** (0.0122)	0.0188 (0.0191)	0.0396*** (0.0147)
Male	-0.0095 (0.0175)	0.0116 (0.0265)	
Non-white	-0.0890* (0.0467)	-0.1277* (0.0709)	
Higher/1st degree	0.1540*** (0.0345)	0.1563*** (0.0466)	
HND/A level	0.0810*** (0.0250)	0.0696* (0.1862)	
CSE/O level	0.0860*** (0.0225)	0.0923*** (0.0327)	
Cut point 1	0.0192 (0.1233)	-0.0277*** (0.2265)	
Cut point 2	1.0698*** (0.1235)	1.0528*** (0.2267)	
σ_u^2		0.0686	
Mean c_i			1.1323
Variance c_i			0.3277
Mean α_i			-0.0743
Variance α_i			0.6311
Correlation(α_i,c_i)			-0.3326
Akaike Infomation Criterion	38544.0	37334.3	37275.2

Standard errors are reported in parentheses. Number of individuals used in estimation of all models is 1739. Estimates of year dummies in all models and within means of variables in random effects are not reported.
* significant at 10% ; ** significant at 5% ; *** significant at 1%.

α_i and c_i as fixed effects. It is estimated by MMLE.

To compare magnitudes of the effects across variables and estimates we look at the relative effects (i.e. ratio of coefficients), and the average and median marginal effects reported in tables 9 and 10 for the variables with a coefficient significantly different from zero.[11,12]

Table 9: Average Marginal Effects on Probability of reporting good and poor health for significant variables. Main model.

(a) Good

	I		II		III	
			Correlated Random			
	Pooled	St.Err.	Effects	St.Err.	MMLE	St.Err.
Health in t-1: Good	0.2528	0.0071	0.1883	0.0456	0.1653	0.0080
Health in t-1: Poor	-0.1550	0.0078	-0.1149	0.0637	-0.1403	0.0520
Age	-0.0005	0.0003	-0.0170	0.0055	-0.0089	0.0064
Household size	-0.0282	0.0050	-0.0040	0.0111	-0.0120	0.0054
Number of Kids	0.0233	0.0056	0.0150	0.0149	0.0145	0.0058
Household Income	0.0294	0.0044	0.0067	0.0095	0.0122	0.0045

(b) Poor

	I		II		III	
			Correlated Random			
	Pooled	St.Err.	Effects	St.Err.	MMLE	St.Err.
Health in t-1: Good	-0.1399	0.0046	-0.1057	0.2125	-0.0984	0.1153
Health in t-1: Poor	0.1477	0.0081	0.0968	0.1372	0.1268	0.0947
Age	0.0003	0.0002	0.0105	0.0140	0.0058	0.0117
Household size	0.0173	0.0031	0.0024	0.0072	0.0081	0.0086
Number of Kids	-0.0143	0.0034	-0.0090	0.0171	-0.0095	0.0102
Household Income	-0.0181	0.0027	-0.0040	0.0078	-0.0081	0.0082

[11] These marginal effects are also called partial effects. The marginal effects are averaged (or calculated their median) across the first eight waves of the panel as well as across the values of the covariates for each individual. This means that we first calculate the marginal effect for each individual in the sample at the observed values of the regressors and then we calculate the average (or the median) of them, instead of calculating the marginal effect at the average value of the covariates. We do this in order to obtain summary measures of the marginal effects representative of the situation of the population (see Chamberlain, 1982, pp.1273). Moreover, a measure that substitutes the values of the covariates and especially the individual specific effect α_i with their means (or any other fixed value) ignores any possible correlation between them. This may give the wrong values of the marginal effects representative of the population.

[12] An alternative way to identify and estimate the marginal effects is the approach taken in Chernozhukov et. al. (2010). They show that in a model like ours, with fixed effects, when T is fixed the (average and quantile) marginal effects are not point identified. However they are set identified and they propose a way to estimate bounds on the partial effect. These nonparametric bounds tighten as T grows. The main advantage is that the bounds analysis applies to any T, whereas our bias correction method depends on T not being very small. However, the bounds analysis is only available with discrete covariates for the moment. In contrast, the bias correction methods work well in many examples, including continuous covariates, and they consistently point estimate the identified average effect.

Table 10: Median Marginal Effects on Probability of reporting good and poor health for significant variables.

(a) Good

	I Pooled	II Corr. Random Effects	III MMLE
Health in t-1: Good	0.2536	0.1889	0.1738
Health in t-1: Poor	-0.1555	-0.1175	-0.1544
Age	-0.0004	-0.0162	-0.0080
Household size	-0.0283	-0.0040	-0.0127
Number of Kids	0.0234	0.0151	0.0154
Household Income	0.0296	0.0067	0.0130

(b) Poor

	I Pooled	II Random Effects	III MMLE
Health in t-1: Good	-0.1402	-0.1014	-0.0910
Health in t-1: Poor	0.1484	0.0949	0.1282
Age	0.0002	0.0094	0.0043
Household size	0.0170	0.0023	0.0077
Number of Kids	-0.0140	-0.0086	-0.0089
Household Income	-0.0177	-0.0039	-0.0077

The pooled model exacerbates the state dependence effect due to the lack of permanent unobserved heterogeneity. Though it is not reported, we also estimated the model in (1) and (2) by MLE. As seen in the simulations it is severely biased, estimating much lower state dependence effects and higher effect of the other explanatory variables.

More interesting is the comparison between the correlated random effects and the fixed effects model estimated by MMLE. They are in columns II and III of Tables 8, 9, and 10. The first difference is in the variables that are statistically significant. Table 8 shows that in the MMLE household size, number of kids, and household income have an impact that is statistically different from zero. However, none of them has a significant effect in the random effect estimates. In correspondence, the average marginal effect of those variables increases in absolute value in the MMLE case with respect to the random effects model, especially for household income. With respect to the state dependence effect (effect of h_{it-1}) there are some changes too. The effect of $h_{it-1} = good$ decreases in absolute value when estimating by MMLE, and the effect of $h_{it-1} = poor$ increases. Comparing coefficients in Table 8 we can also see that the effect of $h_{it-1} = poor$ increases proportionally less than the effect of the other relevant explanatory variables. In the random effects specification the ratio of the coefficient of $\mathbf{1}\left(h_{i,t-1} = poor\right)$ to the coefficient of 'Household income' is around 17, whereas in the MMLE that ratio is 11. In any case, this partial increase in the effect of state dependence and of the effect of the explanatory variables is remarkable because the model in column III allows for more permanent unobserved heterogeneity and more flexibly than in column II.[13]

Moreover, many of those differences in the estimated effects of the explanatory variables between the correlated random effects model and the fixed effects model estimated by MMLE are statistically significant. As is known, if the restrictions imposed by the correlated random effects model are correct its estimates are more precise (i.e. efficient) than the estimates of the fixed effects model (even after the modification of the MLE), though both are consistent. Given this, we have used a Hausman type test to see if those important differences are only due to the more imprecise estimates in columns III. We have made the test over the Average Marginal effects instead of the parameters in table 8 for two reasons. First, Marginal Effects (including their average), and not the parameters in equations (1) and (2), are usually the parameters of interest in nonlinear models. Second, the average marginal effects do not suffer the different scales problem that makes magnitudes in columns II and III of Table 8 not directly comparable and not directly interpretable. The average marginal effects of both models are well defined within the same scale, as any other marginal effect over choice probabilities, and their magnitude has the same clear interpretation. If we were primarily interested in a single average marginal effect, like the effect of $h_{i,t-1} = good$ over the probability of $h_{i,t} = good$, we could use a t-statistic that ignores the others. Doing this for all the average marginal effects we reject at 5% the null hypothesis that both estimates are the same for four variables. Doing a joint test we also reject the null hypothesis that the correlated random effects estimates and the fixed effects MMLE estimates are the same, therefore rejecting, the restrictions imposed in the correlated random effects model.[14]

The previous two paragraphs are a clear indication that ignoring the added dimension of heterogeneity and the flexibility in the distribution of the fixed effects matters when estimating the model and the marginal effects of variables. It is not only a matter of the amount of heterogeneity but also a matter of the other restrictions being imposed on the model in column II.

Besides the formal test of random effects *versus* fixed effects, we look at the unobserved heterogeneity both in the linear index equation and in the cut point shift. Figure 1 displays the estimated distribution (histogram) of both fixed effects in the population, and both exhibit large variation. The average for α_i is -0.074 and for c_i is 1.13. The

[13]Recall that permanent unobserved heterogeneity, state dependence and persistence in observable variables are alternative explanations of the observed high persistence in h_{it}.

[14]In the Hausman test we have used the Var-Cov of the Fixed Effects estimates only, instead of subtracting from it the Var-Cov of the Random Effects. We do this in order to avoid the difference not being a positive definite matrix due to the use of different estimates of the variance of the errors. This represents a lower bound for this test and a rejection here will also be a rejection when using the well defined difference in the var-cov matrices.

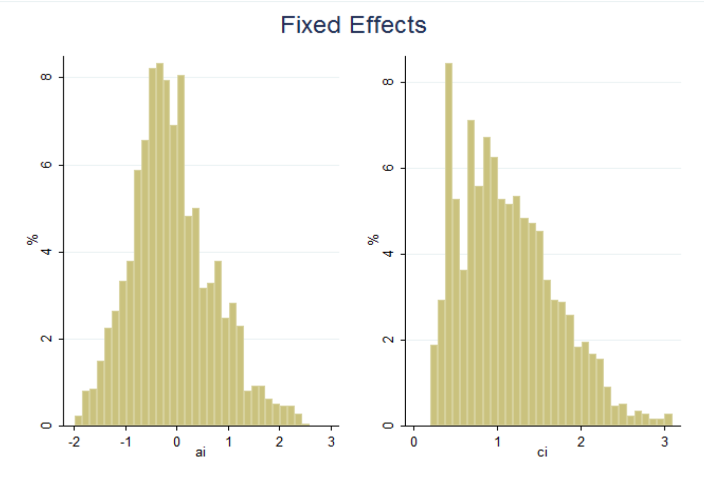

Figure 1: Distribution (histogram) of the fixed effects from MML estimates.

standard deviations of these distributions are 0.79 and 0.57, respectively. In the random effects specification α_i is the compound equation (4) that includes a linear relation to some observables and an additive unobserved term that is assumed to follow a normal distribution. Given the estimates of the parameters of equation (4), the estimated average for α_i in the random effects model is 1.41, and its standard deviation is 0.9626. With respect to the heterogeneity on the cut points, the average of $-c_i$, the first cut point, is -1.13 and the estimate of the first cut point in the random effects specification is -0.03. Also, as can be seen in the right panel of figure 1 and has been said, there is large variation in c_i among individuals that is ignored by the random effects model estimated. Moreover, a test rejects normality of the distribution of α_i at 1%.[15] Finally, the correlation between α_i and c_i is -0.33, so there are rich interactions between both fixed effects forming a joint distribution that is not the simple combination of their marginal distributions.

Focusing on the MML estimates, we find evidence of strong positive state dependence. With respect to socioeconomic variables we find that aging and household size have a small but significant negative effect on SAH. Household income and number of kids have a small but significant positive marginal effect on SAH. Number of kids has the biggest effect of all the x variables.

With respect to how the models fit the observed data, in addition to the information criteria (AIC) reported in Table 8 some predictions of the estimated models and their sample counterparts are in Table 11. Overall the MMLE model fits the data better, because its predictions are closer to the actually observed proportions in the sample. Likewise, the MMLE predictions capture better the inverted-U shape of the proportion of reporting excellent or good health as we look at people with higher number of children, and the slope in the increasing patter when looking at people with higher income.[16]

[15] c_i can not be normal by definition since it is restricted to be positive.

[16] Note that we are not controlling for any other observable characteristics. Thus, there may be other

Table 11: Sample vs. predicted proportions of SAH (in %)

Panel A: Total proportions.

	Poor or very poor	Fair	Excellent or good
Sample	16	31	53
Predicted MMLE	15	32	53
Predicted CRE	12	31	57
Predicted Pooled	14	29	57

Panel B: Proportions of people reporting Excellent or good SAH.

		Predicted		
	Sample	MMLE	CRE	Pooled
By number of Kids				
0	52	53	57	56
1	55	54	55	57
2	58	56	57	60
3+	50	51	54	58
By income quartiles				
1st quartile	47	50	54	54
2nd quartile	51	52	56	56
3rd quartile	56	55	58	59
4th quartile	58	57	59	59

In addition to considering the average and median marginal effects reported in tables 9 and 10, we look at how many individuals have a significant marginal effect in the sample, given their particular situation and unobserved characteristics. Table 12 presents the proportion of individuals with significant (at 10%) marginal effects over the probability of reporting good and bad health, for the same variables as in table 9. Notice that although the average marginal effects are significant, there is a great deal of heterogeneity; for around half the population, the marginal effects over the probability of reporting good health is not significantly different from zero for many of these variables.

4.2 Estimates of additional specifications

4.2.1 Model with health measures

As explained in subsection 2.2, we add variables that contain information on objective health problems to provide further evidence of the robustness and importance of state dependence in SAH. Table 13 presents the estimates of this model by MMLE, and table 14 contains the corresponding average marginal effects. Of the three significant socioeconomic variables in the main model only number of kids remains significantly different

differences between people with different number of children (or different income) that can reinforce or cancel the effect of it on average. Therefore these numbers can not be interpreted as the effect of the number of children (nor the effect of income).

Table 12: Proportion of individuals with marginal effects (on the probability of reporting good and poor) that are significantly different from zero at 10%.

Variable	Proportion	
	Good	Poor
Health in t-1: Good	60.44%	12.25%
Health in t-1: Poor	55.43%	34.50%
Age	22.71%	2.53%
Household size	37.21%	11.44%
Number of Kids	41.81%	12.65%
Household Income	44.85%	15.35%

from zero (at 10%). Most of the objective health measures have the biggest effect over SAH, all with the expected signs. The second variables with higher impact are the two indicators of h_{it-1}.Thus, even after including objective health measures we find evidence of strong positive state dependence here, though it is less than in the main model. The variance of the unobserved heterogeneity is even higher in both α_i and c_i than in the main model.

4.2.2 Linear versus quadratic effect of age

Halliday (2008) found, based on AIC, that a quadratic function of age was only weakly preferred to the linear model and that there was not much lost with a linear model in age. We have estimated model III in table 8 excluding age^2 as an explanatory variable, and in our case the fit is worse because the effect of age increase more than linearly at olger ages. Also, when introducing the quadratic term, the AIC changes much more than in Halliday (2008). Here in the linear model AIC is 37373.4 and in the quadratic model is 37275.2, almost a hundred points smaller.

5 Conclusion

In this paper we have considered the estimation of a dynamic ordered probit of a self-assessed health status with two fixed effects: one in the linear index equation and one in the cut points. The inclusion of two fixed effects, instead of only one as is usual, is motivated by the potential existence of two sources of heterogeneity: unobserved health status and reporting behavior. Even though we cannot separately identify these two sources of heterogeneity we robustly controll for them by using two fixed effects. Based on our best estimates, the two fixed effects exhibit important variation and it is relevant to account for both when estimating the effect of other variables. Our estimates also show that state dependence is large and significant even after controlling for unobserved heterogeneity and

Table 13: Estimates, health indicators added.

Variables	Correlated Random Effects	MMLE
Health in t-1: Good	0.4191*** (0.0337)	0.3696*** (0.0226)
Health in t-1: Poor	-0.1830*** (0.0401)	-0.2784*** (0.0296)
Age	0.0262 (0.0324)	-0.0215 (0.0282)
Age square	-0.0005*** (0.0002)	-0.0003*** (0.0001)
Married	0.0974 (0.1215)	0.0350 (0.0672)
Separated/Divorced	-0.0177 (0.1547)	0.0340 (0.0817)
Widowed	0.1601 (0.2087)	0.0474 (0.1110)
Household size	-0.0181 (0.0359)	-0.0127 (0.0206)
Number of Kids	0.0667 (0.0444)	0.0387* (0.0213)
Household Income	0.0051 (0.0312)	0.0112 (0.0177)
Self employed	-0.0941 (0.1073)	0.0216 (0.0660)
In paid employment	0.1042 (0.0665)	0.1069** (0.0425)
Unemployed	0.1311 (0.0956)	0.0946 (0.0680)
Retired	0.1089 (0.1110)	0.1104* (0.0651)
Long term sick or disa.	-0.1893 (0.1231)	-0.2562*** (0.0707)
Health problems	-0.6808*** (0.0470)	-0.7759*** (0.0334)
Health limits daily acti.	-0.6435*** (0.0465)	-0.6865*** (0.0299)
Health limits work	-0.4956*** (0.0468)	-0.4854*** (0.0306)
Hospital days	-0.0331*** (0.0029)	-0.0350*** (0.0008)
Cut point 1	-0.9318*** (0.2651)	
Cut point 2	0.2788 (0.2647)	
σ_u^2	0.0489	
Mean c_i		1.2775
Variance c_i		0.3942
Mean α_i		2.7760
Variance α_i		1.4170
Correlation(α_i,c_i)		-0.0551
Akaike Infomation Criterion	27688.2	27310.7

Standard errors are reported in parentheses. Number of individuals used in estimation of all models is 1437. Estimates of year dummies in all models, constant variables and within means of variables in random effects are not reported.
* significant at 10% ; ** significant at 5% ; *** significant at 1%.

Table 14: Average Marginal Effects health for significant variables. Model with health indicators added.

(a) Good

	Correlated Random Effects	St.Err.	MMLE	St.Err.
Health in t-1: Good	0.1416	0.0117	0.1122	0.0074
Health in t-1: Poor	-0.0610	0.0134	-0.0832	0.0223
Age	-0.0061	0.0087	-0.0135	0.0080
Number of Kids	0.0213	0.0141	0.0109	0.0060
In paid employment	0.0336	0.0215	0.0306	0.0122
Retired	0.0352	0.0358	0.0316	0.0185
Long term sick or disa.	-0.0610	0.0396	-0.0729	0.0223
Health problems	-0.2250	0.0171	-0.2277	0.0480
Health limits daily acti.	-0.2169	0.0167	-0.2045	0.0340
Health limits work	-0.1666	0.0162	-0.1439	0.0141
Hospital days	-0.0106	0.0009	-0.0099	0.0003

(b) Poor

	Correlated Random Effects	St.Err.	MMLE	St.Err.
Health in t-1: Good	-0.0780	0.0159	-0.0675	0.0877
Health in t-1: Poor	0.0434	0.0119	0.0650	0.0657
Age	0.0038	0.0052	0.0088	0.0161
Number of Kids	-0.0122	0.0081	-0.0070	0.0089
In paid employment	-0.0199	0.0133	-0.0201	0.0247
Retired	-0.0208	0.0213	-0.0207	0.0266
Long term sick or disa.	0.0404	0.0280	0.0547	0.0570
Health problems	0.1083	0.0239	0.1216	0.1667
Health limits daily acti.	0.1435	0.0264	0.1501	0.1630
Health limits work	0.1041	0.0209	0.0994	0.1136
Hospital days	0.0063	0.0012	0.0065	0.0075

some forms of objective health measures. The comparison with random effects estimates previously used shows that it matters to flexibly account for more permanent unobserved heterogeneity.

The recent literature in bias-adjusted methods of estimation of nonlinear panel data models with fixed effects has produced several potentially equivalent estimators. We find that the a priori the most directly applicable correction to our model, which is the HS estimator proposed in Bester and Hansen (2009), still has significant biases in our sample size. This lead us to consider the Modified MLE proposed in Carro (2007). We derive the expression of the MMLE for our model, conduct Monte Carlo experiments to evaluate its finite sample properties, and compare it with the HS. The MMLE has a negligible bias in our sample size. The Monte Carlo experiments contribute to the literature on bias-adjusted methods of estimation nonlinear panel data models by showing how well two of the proposed methods work for a specific model and sample size. This information will be useful for other applications when choosing among the several correction methods existing in the literature.

References

[1] Arellano, M. and J. Hahn (2006): "A likelihood-based approximate solution to the incidental parameter problemin dynamic nonlinear models with multiple effects", *unpublished manuscript.*

[2] Arellano, M. and J. Hahn (2007): "Understanding Bias in Nonlinear Panel Models: Some Recent Developments". in *Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress*, Volume 3, edited by Richard Blundell, Whitney Newey, and Torsten Persson. Cambridge University Press.

[3] Bester, C. A. and C. Hansen (2009): "A Penalty Function Approach to Bias Reduction in Non-linear Panel Models with Fixed Effects". *Journal of Business & Economic Statistic,* 27 (2):131-148

[4] Carro, J. M. (2007) "Estimating dynamic panel data discrete choice models with fixed effects". *Journal of Econometrics,* 140 (2007):503-528

[5] Chamberlain, G. (1984): "Panel Data", in Griliches, Z. and M.D. Intriligator (eds.) *Handbook of Econometrics*, vol. 2, Elsevier Science, Amsterdam.

[6] Chernozhukov, V., I. Fernandez-Val, J. Hahn, and W. Newey (2010): "Average and Quantile Effects in Nonseparable Panel Models", mimeo, MIT Department of Economics.

[7] Contoyannis, P., A. M. Jones and N. Rice (2004): "The Dynamics of Health in the British Household Panel Survey" *Journal of Applied Econometrics,* 19: 473-503

[8] Fernandez-Val, I. (2009): "Fixed effects estimation of structural parameters and marginal effects in panel probit models ", *Journal of Econometrics,* 150 (2009):71-85.

[9] Greene, W. H. and D. A. Henshen (2008): "Modeling Ordered Choices: A Primer and Recent Developments", Available at SSRN: http://ssrn.com/abstract=1213093.

[10] Hahn, J. and G. Kuersteiner (2004): "Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects", mimeo.

[11] Hahn, J. and W. Newey (2004): "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models", *Econometrica,* 72(4): 1295-1319.

[12] Halliday, T. J. (2008): "Heterogeneity, state dependence and health". *Econometrics Journal,* 11: 499-516

[13] Jones, A. M. (2007): "Panel Data Methods and Applications to Health Economics", to appear in *The Palgrave Handbook of Econometrics Volume II: Applied Econometrics*, edited by Terence C. Mills and Kerry Patterson. Basingstoke: Palgrave MacMillan.

[14] Jones, A. M. and S. Schurer (2009): "How does Heterogeneity Shape the Socioeconomic Gradient in Health Satisfaction?". *Journal of Applied Econometrics,* published online, DOI: 10.1002/jae.1134.

[15] Lindeboom, M. and E. van Doorslaer (2004):"Cut-point shift and index shift in self-reported health" *Journal of Health Economics*, 23: 1083-1099

[16] Mundlak, Y. (1978): "On the pooling of time series and cross-section data", *Econometrica,* 46(1): 69-85.

[17] Rilstone, P., V.K. Srivastava, and A. Ullah (1996): "The second-order bias and mean squared error of nonlinear estimators", *Journal of Econometrics*, 75: 369-395

[18] van Doorslaer, E., Jones, A.M. and Koolman, X. (2004). "Explaining income-related inequalities in doctor utilisation in Europe" *Health Economics*, 13: 629-647

[19] Wooldridge, J. (2005): "Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity ", *Journal of Applied Econometrics,* 20: 39-54.

A Appendix: Reduction of the order of the bias

In this appendix we show that the modified score presented above corrects the first order asymptotic bias of the original score. The algebra is somewhat tedious because of the many terms, but the idea is clear. We first expand the score of the MLE around the true value of the fixed effects and make some calculations and substitutions on it to obtain the leading term of the bias of the MLE's score. Then we show that the modification in the MMLE's score, equation (10), is subtracting that leading bias term from the score. This follows Carro (2007), and is adapted to our model with two fixed effects.

The notation used is the same as in section 3.2: $\gamma= (\beta, \rho_1, \rho_{-1})$ and $\eta_i = (\alpha_i, c_i)$; we denote partial derivatives by the letter d; bold letters are used to denote vectors; $\mathbf{d}_{\eta i} \equiv \frac{\partial l_i(\gamma,\eta_i)}{\partial \eta_i}$, $\mathbf{d}_{\gamma i} \equiv \frac{\partial l_i(\gamma,\eta_i)}{\partial \gamma}$, $\mathbf{d}_{\gamma ci} = \frac{\partial^2 l_i}{\partial\gamma\partial c_i}$, $d_{\alpha\alpha i} = \frac{\partial^2 l_i}{\partial \alpha_i^2}$, $\mathbf{d}_{\gamma\alpha ci} = \frac{\partial^3 l_i}{\partial\gamma\partial c_i\partial\alpha_i}$, and so on; the derivatives evaluated at the true values of the parameters are represented by including a 0 in the sub-index (e.g. $d_{\eta i0} = d_{\eta i}(\gamma_0, \eta_{i0})$).

A.1 Deriving the leading term of the bias of the score in the MLE

We start by deriving the first term of the bias in the score of the original unmodified concentrated log-likelihood. Expanding this score around η_{i0}, and evaluating it at γ_0 we get:

$$\begin{aligned}\mathbf{d}_{\gamma i}(\gamma_0, \eta_i(\gamma_0)) = \mathbf{d}_{\gamma i0} + \; & d_{\gamma\alpha i0}(\hat{\alpha}_i(\gamma_0) - \alpha_{i0}) \\ & + \mathbf{d}_{\gamma ci0}(\hat{c}_i(\gamma_0) - c_{i0}) \\ & + \frac{1}{2}\mathbf{d}_{\gamma\alpha\alpha i0}(\hat{\alpha}_i(\gamma_0) - \alpha_{i0})^2 + \frac{1}{2}\mathbf{d}_{\gamma cci0}(\hat{c}_i(\gamma_0) - c_{i0})^2 \\ & + \mathbf{d}_{\gamma aci0}(\hat{\alpha}_i(\gamma_0) - \alpha_{i0})(\hat{c}_i(\gamma_0) - c_{i0}) + O_p(T^{-1/2}) + \dots\end{aligned} \tag{A1}$$

This equation clearly shows that the score evaluated at the true value γ_0 differs from the value of the score we want to obtain, $\mathbf{d}_{\gamma i0} = \mathbf{d}_{\gamma i}(\gamma_0, \eta_{i0})$, as much as $\hat{\alpha}_i(\gamma_0)$ and $\hat{c}_i(\gamma_0)$ differ from α_{i0} and c_{i0}. This is the source of the incidental parameters problem.

Now we need expressions for $(\hat{\alpha}_i(\gamma_0) - \alpha_{i0})$ and $(\hat{c}_i(\gamma_0) - c_{i0})$, for which we do asymptotic expansions, following Rilstone, Srivastava and Ullah (1996):

$$(\hat{\alpha}_i(\gamma_0) - \alpha_{i0}) = b^{\alpha}_{-1/2} + b^{\alpha}_{-1} + O_p(T^{-3/2}) \tag{A2}$$

$$(\hat{c}_i(\gamma_0) - c_{i0}) = b^{c}_{-1/2} + b^{c}_{-1} + O_p(T^{-3/2}) \tag{A3}$$

where $b^{\alpha}_{-1/2}$ and $b^{c}_{-1/2}$ are the elements of the vector $\mathbf{b}_{-1/2}$, and b^{α}_{-1} and b^{c}_{-1} are the elements of the vector $\mathbf{b}_{-1}$, which are determined as follows:

$$\begin{aligned}\mathbf{b}_{-1/2} &= -Q^{-1}R \\ \mathbf{b}_{-1} &= -Q^{-1}S\,\mathbf{b}_{-1/2} - \frac{1}{2}Q^{-1}U\,(\mathbf{b}_{-1/2} \otimes \mathbf{b}_{-1/2}) \\ R &= \frac{1}{T}\begin{pmatrix} d_{\alpha i0} \\ d_{ci0} \end{pmatrix} \\ Q &= E(\nabla R) \\ S &= \nabla R - Q \\ U &= E(\nabla^2 Q)\end{aligned}$$

From the above expressions we obtain:

$$b^{\alpha}_{-1/2} = \frac{\frac{1}{T} d_{ci0} E\left(\frac{1}{T}\, d_{c\alpha i0}\right) - \frac{1}{T} d_{ai0} E\left(\frac{1}{T}\, d_{cci0}\right)}{E\left(\frac{1}{T}\, d_{\alpha\alpha i0}\right) E\left(\frac{1}{T}\, d_{cci0}\right) - E\left(\frac{1}{T}\, d_{c\alpha i0}\right)^2} \tag{A4}$$

$$b^{c}_{-1/2} = \frac{\frac{1}{T} d_{ai0} E\left(\frac{1}{T}\, d_{c\alpha i0}\right) - \frac{1}{T} d_{ci0} E\left(\frac{1}{T}\, d_{\alpha\alpha i0}\right)}{E\left(\frac{1}{T}\, d_{\alpha\alpha i0}\right) E\left(\frac{1}{T}\, d_{cci0}\right) - E\left(\frac{1}{T}\, d_{c\alpha i0}\right)^2} \tag{A5}$$

It is also useful to obtain:

$$(\hat{\alpha}_i(\gamma_0) - \alpha_{i0})^2 = (b^{a}_{-1/2})^2 + O_p(T^{-3/2}) \tag{A6}$$

$$(\hat{c}_i(\gamma_0) - c_{i0})^2 = (b^{c}_{-1/2})^2 + O_p(T^{-3/2}) \tag{A7}$$

$$(\hat{\alpha}_i(\gamma_0) - \alpha_{i0})\,(\hat{c}_i(\gamma_0) - c_{i0}) = b^{a}_{-1/2}\, b^{c}_{-1/2} + O_p(T^{-3/2}) \tag{A8}$$

With respect to the squares of $b^{\alpha}_{-1/2}$ and $b^{c}_{-1/2}$, we get:

$$(b^{\alpha}_{-1/2})^2 = \frac{\left(\frac{1}{T} d_{ai0}\right)^2 E\left(\frac{1}{T}\, d_{cci0}\right)^2 + \left(\frac{1}{T} d_{ci0}\right)^2 E\left(\frac{1}{T}\, d_{c\alpha i0}\right)^2 - 2\, \frac{1}{T} d_{ai0} \frac{1}{T} d_{ci0} E\left(\frac{1}{T}\, d_{c\alpha i0}\right) E\left(\frac{1}{T}\, d_{cci0}\right)}{\left(E\left(\frac{1}{T}\, d_{\alpha\alpha i0}\right) E\left(\frac{1}{T}\, d_{cci0}\right) - E\left(\frac{1}{T}\, d_{c\alpha i0}\right)^2\right)^2}$$

$$(b^{c}_{-1/2})^2 = \frac{\left(\frac{1}{T} d_{ci0}\right)^2 E\left(\frac{1}{T}\, d_{\alpha\alpha i0}\right)^2 + \left(\frac{1}{T} d_{ai0}\right)^2 E\left(\frac{1}{T}\, d_{c\alpha i0}\right)^2 - 2\, \frac{1}{T} d_{ai0} \frac{1}{T} d_{ci0} E\left(\frac{1}{T}\, d_{\alpha\alpha i0}\right) E\left(\frac{1}{T}\, d_{c\alpha i0}\right)}{\left(E\left(\frac{1}{T}\, d_{\alpha\alpha i0}\right) E\left(\frac{1}{T}\, d_{cci0}\right) - E\left(\frac{1}{T}\, d_{c\alpha i0}\right)^2\right)^2}$$

Substituting by expectations, and using the information matrix identity ($E(d_{c\alpha i}) = -E(d_{ai} d_{ci})$), we get:

$$(b^{\alpha}_{-1/2})^2 = -\frac{1}{T} \frac{E\left(\frac{1}{T}\, d_{cci0}\right)}{E\left(\frac{1}{T}\, d_{\alpha\alpha i0}\right) E\left(\frac{1}{T}\, d_{cci0}\right) - E\left(\frac{1}{T}\, d_{c\alpha i0}\right)^2} + O_p(T^{-3/2}) \tag{A9}$$

$$(b^{c}_{-1/2})^2 = -\frac{1}{T} \frac{E\left(\frac{1}{T}\, d_{\alpha\alpha i0}\right)}{E\left(\frac{1}{T}\, d_{\alpha\alpha i0}\right) E\left(\frac{1}{T}\, d_{cci0}\right) - E\left(\frac{1}{T}\, d_{c\alpha i0}\right)^2} + O_p(T^{-3/2}) \tag{A10}$$

Following the same procedure for the cross-product, we get:

$$b^{\alpha}_{-1/2}\, b^{c}_{-1/2} = \frac{1}{T} \frac{E\left(\frac{1}{T}\, d_{c\alpha i0}\right)}{E\left(\frac{1}{T}\, d_{\alpha\alpha i0}\right) E\left(\frac{1}{T}\, d_{cci0}\right) - E\left(\frac{1}{T}\, d_{c\alpha i0}\right)^2} + O_p(T^{-3/2}) \tag{A11}$$

With respect to b^{α}_{-1} and b^{c}_{-1}, we follow the same procedure (replace by expectations and use the information matrix identity) to get:

$$
\begin{aligned}
b^{\alpha}_{-1} = {} & \frac{1}{2\,T}\,\frac{1}{\left(E\left(\frac{1}{T}\,d_{\alpha\alpha i0}\right)E\left(\frac{1}{T}\,d_{cci0}\right)-E\left(\frac{1}{T}\,d_{c\alpha i0}\right)^2\right)^2} \qquad \text{(A12)}\\
& \Bigg\{2\,E\left(\frac{1}{T}\,d_{c\alpha i0}\right)^2\left(E\left(\frac{1}{T}\,d_{\alpha cci0}\right)+E\left(\frac{1}{T}\,d_{ai0}d_{cci0}\right)+E\left(\frac{1}{T}\,d_{ci0}d_{c\alpha i0}\right)\right)\\
& +E\left(\frac{1}{T}\,d_{cci0}\right)^2\left[E\left(\frac{1}{T}\,d_{a\alpha\alpha i0}\right)+2E\left(\frac{1}{T}\,d_{ai0}d_{\alpha\alpha i0}\right)\right]\\
& +E\left(\frac{1}{T}\,d_{\alpha\alpha i0}\right)E\left(\frac{1}{T}\,d_{cci0}\right)\left[E\left(\frac{1}{T}\,d_{\alpha cci0}\right)+2E\left(\frac{1}{T}\,d_{ci0}d_{c\alpha i0}\right)\right]\\
& -E\left(\frac{1}{T}\,d_{c\alpha i0}\right)E\left(\frac{1}{T}\,d_{\alpha\alpha i0}\right)\left[E\left(\frac{1}{T}\,d_{ccci0}\right)+2E\left(\frac{1}{T}\,d_{ci0}d_{cci0}\right)\right]\\
& -E\left(\frac{1}{T}\,d_{c\alpha i0}\right)E\left(\frac{1}{T}\,d_{cci0}\right)\left[3E\left(\frac{1}{T}\,d_{\alpha\alpha ci0}\right)+4E\left(\frac{1}{T}\,d_{ai0}d_{c\alpha i0}\right)+2E\left(\frac{1}{T}\,d_{ci0}d_{\alpha\alpha i0}\right)\right]\Bigg\}\\
& +O_p(T^{-3/2})\\
b^{c}_{-1} = {} & \frac{1}{2\,T}\,\frac{1}{\left(E\left(\frac{1}{T}\,d_{\alpha\alpha i0}\right)E\left(\frac{1}{T}\,d_{cci0}\right)-E\left(\frac{1}{T}\,d_{c\alpha i0}\right)^2\right)^2} \qquad \text{(A13)}\\
& \Bigg\{2\,E\left(\frac{1}{T}\,d_{c\alpha i0}\right)^2\left[E\left(\frac{1}{T}\,d_{\alpha\alpha ci0}\right)+E\left(\frac{1}{T}\,d_{ci0}d_{\alpha\alpha i0}\right)+E\left(\frac{1}{T}\,d_{ai0}d_{c\alpha i0}\right)\right]\\
& +E\left(\frac{1}{T}\,d_{\alpha\alpha i0}\right)^2\left[E\left(\frac{1}{T}\,d_{ccci0}\right)+2E\left(\frac{1}{T}\,d_{ci0}d_{cci0}\right)\right]\\
& +E\left(\frac{1}{T}\,d_{\alpha\alpha i0}\right)E\left(\frac{1}{T}\,d_{cci0}\right)\left[E\left(\frac{1}{T}\,d_{\alpha\alpha ci0}\right)+2E\left(\frac{1}{T}\,d_{ai0}d_{c\alpha i0}\right)\right]\\
& -E\left(\frac{1}{T}\,d_{c\alpha i0}\right)E\left(\frac{1}{T}\,d_{cci0}\right)\left[E\left(\frac{1}{T}\,d_{a\alpha\alpha i0}\right)+2E\left(\frac{1}{T}\,d_{ai0}d_{\alpha\alpha i0}\right)\right]\\
& -E\left(\frac{1}{T}\,d_{c\alpha i0}\right)E\left(\frac{1}{T}\,d_{\alpha\alpha i0}\right)\left[3E\left(\frac{1}{T}\,d_{\alpha cci0}\right)+4E\left(\frac{1}{T}\,d_{ci0}d_{c\alpha i0}\right)+2E\left(\frac{1}{T}\,d_{ai0}d_{cci0}\right)\right]\Bigg\}\\
& +O_p(T^{-3/2}) \qquad \text{(A14)}
\end{aligned}
$$

Introducing all these expressions in (A1), and taking expectations, we get:

$$
\begin{aligned}
&E(d_{\gamma i}(\gamma_0, \hat{\eta}_i(\gamma_0))) = \qquad\qquad (\mathrm{A15})\\
&\frac{E\left(\frac{1}{T}\,\mathbf{d}_{\gamma\alpha i0} d_{ci0}\right) E\left(\frac{1}{T}\, d_{c\alpha i0}\right) - E\left(\frac{1}{T}\,\mathbf{d}_{\gamma\alpha i0} d_{ai0}\right) E\left(\frac{1}{T}\, d_{cci0}\right)}{E\left(\frac{1}{T}\, d_{\alpha\alpha i0}\right) E\left(\frac{1}{T}\, d_{cci0}\right) - E\left(\frac{1}{T}\, d_{c\alpha i0}\right)^2}\\
&+ \frac{1}{2}\, \frac{E\left(\frac{1}{T}\,\mathbf{d}_{\gamma ai0}\right)}{\left(E\left(\frac{1}{T}\, d_{\alpha\alpha i0}\right) E\left(\frac{1}{T}\, d_{cci0}\right) - E\left(\frac{1}{T}\, d_{c\alpha i0}\right)^2\right)^2}\\
&\left\{ 2\,E\left(\frac{1}{T}\, d_{c\alpha i0}\right)^2 \left(E\left(\frac{1}{T}\, d_{\alpha cci0}\right) + E\left(\frac{1}{T}\, d_{ai0} d_{cci0}\right) + E\left(\frac{1}{T}\, d_{ci0} d_{c\alpha i0}\right)\right)\right.\\
&+ E\left(\frac{1}{T}\, d_{cci0}\right)^2 \left[E\left(\frac{1}{T}\, d_{a\alpha\alpha i0}\right) + 2E\left(\frac{1}{T}\, d_{ai0} d_{\alpha\alpha i0}\right)\right]\\
&+ E\left(\frac{1}{T}\, d_{\alpha\alpha i0}\right) E\left(\frac{1}{T}\, d_{cci0}\right) \left[E\left(\frac{1}{T}\, d_{\alpha cci0}\right) + 2E\left(\frac{1}{T}\, d_{ci0} d_{c\alpha i0}\right)\right]\\
&- E\left(\frac{1}{T}\, d_{c\alpha i0}\right) E\left(\frac{1}{T}\, d_{\alpha\alpha i0}\right) \left[E\left(\frac{1}{T}\, d_{ccci0}\right) + 2E\left(\frac{1}{T}\, d_{ci0} d_{cci0}\right)\right]\\
&\left. -E\left(\frac{1}{T}\, d_{c\alpha i0}\right) E\left(\frac{1}{T}\, d_{cci0}\right) \left[3E\left(\frac{1}{T}\, d_{\alpha\alpha ci0}\right) + 4E\left(\frac{1}{T}\, d_{ai0} d_{c\alpha i0}\right) + 2E\left(\frac{1}{T}\, d_{ci0} d_{\alpha\alpha i0}\right)\right]\right\}\\
&+ \frac{E\left(\frac{1}{T}\,\mathbf{d}_{\gamma ci0} d_{ai0}\right) E\left(\frac{1}{T}\, d_{c\alpha i0}\right) - E\left(\frac{1}{T}\,\mathbf{d}_{\gamma ci0} d_{ci0}\right) E\left(\frac{1}{T}\, d_{\alpha\alpha i0}\right)}{E\left(\frac{1}{T}\, d_{\alpha\alpha i0}\right) E\left(\frac{1}{T}\, d_{cci0}\right) - E\left(\frac{1}{T}\, d_{c\alpha i0}\right)^2}\\
&+ \frac{1}{2}\, \frac{E\left(\frac{1}{T}\,\mathbf{d}_{\gamma ci0}\right)}{\left(E\left(\frac{1}{T}\, d_{\alpha\alpha i0}\right) E\left(\frac{1}{T}\, d_{cci0}\right) - E\left(\frac{1}{T}\, d_{c\alpha i0}\right)^2\right)^2}\\
&\left\{ 2\,E\left(\frac{1}{T}\, d_{c\alpha i0}\right)^2 \left[E\left(\frac{1}{T}\, d_{\alpha\alpha ci0}\right) + E\left(\frac{1}{T}\, d_{ci0} d_{\alpha\alpha i0}\right) + E\left(\frac{1}{T}\, d_{ai0} d_{c\alpha i0}\right)\right]\right.\\
&+ E\left(\frac{1}{T}\, d_{\alpha\alpha i0}\right)^2 \left[E\left(\frac{1}{T}\, d_{ccci0}\right) + 2E\left(\frac{1}{T}\, d_{ci0} d_{cci0}\right)\right]\\
&+ E\left(\frac{1}{T}\, d_{\alpha\alpha i0}\right) E\left(\frac{1}{T}\, d_{cci0}\right) \left[E\left(\frac{1}{T}\, d_{\alpha\alpha ci0}\right) + 2E\left(\frac{1}{T}\, d_{ai0} d_{c\alpha i0}\right)\right]\\
&- E\left(\frac{1}{T}\, d_{c\alpha i0}\right) E\left(\frac{1}{T}\, d_{cci0}\right) \left[E\left(\frac{1}{T}\, d_{a\alpha\alpha i0}\right) + 2E\left(\frac{1}{T}\, d_{ai0} d_{\alpha\alpha i0}\right)\right]\\
&\left. -E\left(\frac{1}{T}\, d_{c\alpha i0}\right) E\left(\frac{1}{T}\, d_{\alpha\alpha i0}\right) \left[3E\left(\frac{1}{T}\, d_{\alpha cci0}\right) + 4E\left(\frac{1}{T}\, d_{ci0} d_{c\alpha i0}\right) + 2E\left(\frac{1}{T}\, d_{ai0} d_{cci0}\right)\right]\right\}\\
&+ \frac{1}{E\left(\frac{1}{T}\, d_{\alpha\alpha i0}\right) E\left(\frac{1}{T}\, d_{cci0}\right) - E\left(\frac{1}{T}\, d_{c\alpha i0}\right)^2}\\
&\left[E\left(\frac{1}{T}\,\mathbf{d}_{\gamma aci0}\right) E\left(\frac{1}{T}\, d_{c\alpha i0}\right) - \frac{1}{2}E\left(\frac{1}{T}\,\mathbf{d}_{\gamma\alpha\alpha i0}\right) E\left(\frac{1}{T}\, d_{cci0}\right) - \frac{1}{2}E\left(\frac{1}{T}\,\mathbf{d}_{\gamma cci0}\right) E\left(\frac{1}{T}\, d_{\alpha\alpha i0}\right)\right]\\
&+ O(T^{-1})
\end{aligned}
$$

The remainder of this expression is $O(T^{-1})$ because $O_p(T^{-1/2})$ terms have zero mean. This means that the score of the original concentrated likelihood has a bias of order $O(1)$, whose expression is in the previous formulae.

A.2 Modified Score

The modified score in (10) can be decomposed in three terms, $\mathbf{d}_{\gamma Mi}(\gamma) = A + B + C$, such that:

$$
\begin{aligned}
A &= \mathbf{d}_{\gamma i}(\gamma, \eta_i(\gamma)) && \text{(A16)}\\
B &= -\frac{1}{2}\frac{1}{d_{\alpha\alpha i}d_{cci} - {d_{c\alpha i}}^2} && \text{(A17)}\\
&\left[d_{\alpha\alpha i}\left(\mathbf{d}_{\gamma cci} + d_{\alpha cci}\frac{\partial\hat{\alpha}_i}{\partial\gamma} + d_{ccci}\frac{\partial\hat{c}_i}{\partial\gamma}\right)\right.\\
&+ d_{cci}\left(\mathbf{d}_{\gamma\alpha\alpha i} + d_{\alpha\alpha\alpha i}\frac{\partial\hat{\alpha}_i}{\partial\gamma} + d_{\alpha\alpha ci}\frac{\partial\hat{c}_i}{\partial\gamma}\right)\\
&\left.-2d_{c\alpha i}\left(\mathbf{d}_{\gamma aci} + d_{\alpha\alpha ci}\frac{\partial\hat{\alpha}_i}{\partial\gamma} + d_{\alpha cci}\frac{\partial\hat{c}_i}{\partial\gamma}\right)\right]\\
C &= -\frac{\partial}{\partial a_i}\left(\frac{E(\mathbf{d}_{\gamma ci})E(d_{c\alpha i}) - E(d_{cci})E(\mathbf{d}_{\gamma\alpha i})}{E(d_{\alpha\alpha i})E(d_{cci}) - [E(d_{c\alpha i})]^2}\right)\bigg|_{\eta_i=\eta_i(\gamma)} && \text{(A18)}\\
&-\frac{\partial}{\partial c_i}\left(\frac{E(\mathbf{d}_{\gamma\alpha i})E(d_{c\alpha i}) - E(d_{\alpha\alpha i})E(\mathbf{d}_{\gamma ci})}{E(d_{\alpha\alpha i})E(d_{cci}) - [E(d_{c\alpha i})]^2}\right)\bigg|_{\eta_i=\eta_i(\gamma)}
\end{aligned}
$$

A is the score of the original unmodified concentrated log-likelihood. So, we now analyze B and C.

Part B**.** We first want to derive expression for $\partial\hat{\alpha}_i/\partial\gamma$ and $\partial\hat{c}_i/\partial\gamma$. Differentiating the score of the concentrated log-likelihood, $\mathbf{d}_{\eta i}(\gamma, \eta_i(\gamma))$, with respect to γ we get a system of two equations with two unknowns. Solving for $\partial\hat{\alpha}_i/\partial\gamma$ and $\partial\hat{c}_i/\partial\gamma$ we get:

$$
\frac{\partial\hat{\alpha}_i(\gamma)}{\partial\gamma} = \frac{\mathbf{d}_{\gamma ci}d_{c\alpha i} - d_{cci}\mathbf{d}_{\gamma\alpha i}}{d_{\alpha\alpha i}d_{cci} - d_{c\alpha i}^2} \tag{A19}
$$

$$
\frac{\partial\hat{c}_i(\gamma)}{\partial\gamma} = \frac{\mathbf{d}_{\gamma\alpha i}d_{c\alpha i} - d_{\alpha\alpha i}\mathbf{d}_{\gamma ci}}{d_{\alpha\alpha i}d_{cci} - d_{c\alpha i}^2} \tag{A20}
$$

evaluating at γ_0 and replacing by expectations:

$$
\frac{\partial\hat{\alpha}_i(\gamma_0)}{\partial\gamma} = \frac{E\left(\frac{1}{T}\,\mathbf{d}_{\gamma ci0}\right)E\left(\frac{1}{T}\,d_{c\alpha i0}\right) - E\left(\frac{1}{T}\,d_{cci0}\right)E\left(\frac{1}{T}\,\mathbf{d}_{\gamma\alpha i0}\right)}{E\left(\frac{1}{T}\,d_{\alpha\alpha i0}\right)E\left(\frac{1}{T}\,d_{cci0}\right) - E\left(\frac{1}{T}\,d_{c\alpha i0}\right)^2} + O_p(T^{-\frac{1}{2}}) \tag{A21}
$$

$$
\frac{\partial\hat{c}_i(\gamma_0)}{\partial\gamma} = \frac{E\left(\frac{1}{T}\,\mathbf{d}_{\gamma\alpha i0}\right)E\left(\frac{1}{T}\,d_{c\alpha i0}\right) - E\left(\frac{1}{T}\,d_{\alpha\alpha i0}\right)E\left(\frac{1}{T}\,\mathbf{d}_{\gamma ci0}\right)}{E\left(\frac{1}{T}\,d_{\alpha\alpha i0}\right)E\left(\frac{1}{T}\,d_{cci0}\right) - E\left(\frac{1}{T}\,d_{c\alpha i0}\right)^2} + O_p(T^{-\frac{1}{2}}) \tag{A22}
$$

Introducing in (A17) and rearranging terms:

$$
\begin{aligned}
B = &-\frac{E\left(\frac{1}{T}\,\mathbf{d}_{\gamma ci0}\right)E\left(\frac{1}{T}\,d_{c\alpha i0}\right)-E\left(\frac{1}{T}\,d_{cci0}\right)E\left(\frac{1}{T}\,\mathbf{d}_{\gamma\alpha i0}\right)}{E\left(\frac{1}{T}\,d_{\alpha\alpha i0}\right)E\left(\frac{1}{T}\,d_{cci0}\right)-E\left(\frac{1}{T}\,d_{c\alpha i0}\right)^2} \\
&\frac{d_{\alpha\alpha i}d_{\alpha cci}+d_{cci}d_{\alpha\alpha\alpha i}-2d_{c\alpha i}d_{\alpha\alpha ci}}{2(d_{\alpha\alpha i}d_{cci}-d_{c\alpha i}^2)} \\
&-\frac{d_{\alpha\alpha i}d_{\alpha cci}+d_{cci}d_{\alpha\alpha\alpha i}-2d_{c\alpha i}d_{\alpha\alpha ci}}{2(d_{\alpha\alpha i}d_{cci}-d_{c\alpha i}^2)}\,O_p(T^{-1/2}) \\
&-\frac{E\left(\frac{1}{T}\,\mathbf{d}_{\gamma\alpha i0}\right)E\left(\frac{1}{T}\,d_{c\alpha i0}\right)-E\left(\frac{1}{T}\,d_{\alpha\alpha i0}\right)E\left(\frac{1}{T}\,\mathbf{d}_{\gamma ci0}\right)}{E\left(\frac{1}{T}\,d_{\alpha\alpha i0}\right)E\left(\frac{1}{T}\,d_{cci0}\right)-E\left(\frac{1}{T}\,d_{c\alpha i0}\right)^2} \\
&\frac{d_{cci}d_{\alpha\alpha ci}+d_{\alpha\alpha i}d_{ccci}-2d_{c\alpha i}d_{\alpha cci}}{2(d_{\alpha\alpha i}d_{cci}-d_{c\alpha i}^2)} \\
&-\frac{d_{cci}d_{\alpha\alpha ci}+d_{\alpha\alpha i}d_{ccci}-2d_{c\alpha i}d_{\alpha cci}}{2(d_{\alpha\alpha i}d_{cci}-d_{c\alpha i}^2)}\,O_p(T^{-1/2}) \\
&-\frac{d_{\alpha\alpha i}\mathbf{d}_{\gamma cci}+d_{cci}\mathbf{d}_{\gamma\alpha\alpha i}-2d_{c\alpha i}\mathbf{d}_{\gamma aci}}{2(d_{\alpha\alpha i}d_{cci}-d_{c\alpha i}^2)}
\end{aligned}
\tag{A23}
$$

Evaluating at γ_0, using the fact that $\eta_i(\gamma) = \eta_{i0} + O_p(T^{-1/2})$, adding $1/T^2$ in numerators and denominators and replacing by expectations:

$$
\begin{aligned}
B = &-\frac{1}{2}\frac{1}{\left(E\left(\frac{1}{T}\,d_{\alpha\alpha i0}\right)E\left(\frac{1}{T}\,d_{cci0}\right)-E\left(\frac{1}{T}\,d_{c\alpha i0}\right)^2\right)^2} \\
&\Bigg\{\left[E\left(\frac{1}{T}\,\mathbf{d}_{\gamma ci0}\right)E\left(\frac{1}{T}\,d_{c\alpha i0}\right)-E\left(\frac{1}{T}\,d_{cci0}\right)E\left(\frac{1}{T}\,\mathbf{d}_{\gamma ai0}\right)\right] \\
&\left[E\left(\frac{1}{T}\,d_{\alpha\alpha i0}\right)E\left(\frac{1}{T}\,d_{\alpha cci0}\right)+E\left(\frac{1}{T}\,d_{cci0}\right)E\left(\frac{1}{T}\,d_{a\alpha\alpha i0}\right)-2E\left(\frac{1}{T}\,d_{c\alpha i0}\right)E\left(\frac{1}{T}\,d_{\alpha\alpha ci0}\right)\right] \\
&+\left[E\left(\frac{1}{T}\,\mathbf{d}_{\gamma\alpha i0}\right)E\left(\frac{1}{T}\,d_{c\alpha i0}\right)-E\left(\frac{1}{T}\,d_{\alpha\alpha i0}\right)E\left(\frac{1}{T}\,\mathbf{d}_{\gamma ci0}\right)\right] \\
&\left[E\left(\frac{1}{T}\,d_{cci0}\right)E\left(\frac{1}{T}\,d_{\alpha\alpha ci0}\right)+E\left(\frac{1}{T}\,d_{\alpha\alpha i0}\right)E\left(\frac{1}{T}\,d_{ccci0}\right)-2E\left(\frac{1}{T}\,d_{c\alpha i0}\right)E\left(\frac{1}{T}\,d_{\alpha cci0}\right)\right]\Bigg\} \\
&-\frac{1}{2}\frac{1}{E\left(\frac{1}{T}\,d_{\alpha\alpha i0}\right)E\left(\frac{1}{T}\,d_{cci0}\right)-E\left(\frac{1}{T}\,d_{c\alpha i0}\right)^2} \\
&\left[E\left(\frac{1}{T}\,d_{\alpha\alpha i0}\right)E\left(\frac{1}{T}\,\mathbf{d}_{\gamma cci0}\right)+E\left(\frac{1}{T}\,d_{cci0}\right)E\left(\frac{1}{T}\,\mathbf{d}_{\gamma\alpha\alpha i0}\right)-2E\left(\frac{1}{T}\,d_{c\alpha i0}\right)E\left(\frac{1}{T}\,\mathbf{d}_{\gamma aci0}\right)\right] \qquad \text{(A24)} \\
&+O_p(T^{-1/2}) \qquad \text{(A25)}
\end{aligned}
$$

Finally, taking the expected value of this expression will not change anything, except that the remainder would be $O(T^{-1})$ instead of $O_p(T^{-1/2})$.

Part C. To analyze C, we need the following result:

$$
\frac{\partial}{\partial\alpha_i}E\left(\mathbf{d}_{\gamma ci}\right) = E\left(\mathbf{d}_{\gamma\alpha ci}\right)+E\left(\mathbf{d}_{\gamma ci}d_{\alpha i}\right)
\tag{A26}
$$

This works with other derivatives of expectations as well.

C is the sum of two derivatives, that we call C^α and C^c respectively, evaluated at $\eta_i = \eta_i(\gamma)$. C^α is equal to:

$$
\begin{aligned}
C^{\alpha} &= -\frac{\partial}{\partial a_i}\left(\frac{E(\mathbf{d}_{\gamma ci})E(d_{c\alpha i}) - E(d_{cci})E(\mathbf{d}_{\gamma\alpha i})}{E(d_{\alpha\alpha i})E(d_{cci}) - [E(d_{c\alpha i})]^2}\right) \\
&= -\frac{\frac{\partial}{\partial a_i}\left(E(\mathbf{d}_{\gamma ci})E(d_{c\alpha i}) - E(d_{cci})E(\mathbf{d}_{\gamma\alpha i})\right)}{E(d_{\alpha\alpha i})E(d_{cci}) - [E(d_{c\alpha i})]^2} \\
&+ \frac{\left(E(\mathbf{d}_{\gamma ci})E(d_{c\alpha i}) - E(d_{cci})E(\mathbf{d}_{\gamma\alpha i})\right)\frac{\partial}{\partial a_i}\left(E(d_{\alpha\alpha i})E(d_{cci}) - [E(d_{c\alpha i})]^2\right)}{\left(E(d_{\alpha\alpha i})E(d_{cci}) - [E(d_{c\alpha i})]^2\right)^2}
\end{aligned}
$$

Working with the derivative and using the above result, we get:

$$
\begin{aligned}
C^{a} = &-\frac{1}{E(d_{\alpha\alpha i})E(d_{cci}) - [E(d_{c\alpha i})]^2} \\
&\{E(\mathbf{d}_{\gamma ci})\left[E(d_{\alpha\alpha ci}) + E(d_{c\alpha i}d_{ai})\right] + E(d_{c\alpha i})\left[E(\mathbf{d}_{\gamma aci}) + E(\mathbf{d}_{\gamma ci}d_{ai})\right] \\
&-E(d_{cci})\left[E(\mathbf{d}_{\gamma\alpha\alpha i}) + E(\mathbf{d}_{\gamma\alpha i}d_{ai})\right] - E(\mathbf{d}_{\gamma\alpha i})\left[E(d_{\alpha cci}) + E(d_{cci}d_{ai})\right]\} \\
&+ \frac{E(\mathbf{d}_{\gamma ci})E(d_{c\alpha i}) - E(d_{cci})E(\mathbf{d}_{\gamma\alpha i})}{\left(E(d_{\alpha\alpha i})E(d_{cci}) - [E(d_{c\alpha i})]^2\right)^2} \\
&\{E(d_{\alpha\alpha i})\left[E(d_{\alpha cci}) + E(d_{cci}d_{ai})\right] + E(d_{cci})\left[E(d_{\alpha\alpha\alpha i}) + E(d_{\alpha\alpha i}d_{ai})\right] \\
&-2E(d_{c\alpha i})\left[E(d_{\alpha\alpha ci}) + E(d_{c\alpha i}d_{ai})\right]\}
\end{aligned}
$$

Likewise, for C^c we have:

$$
\begin{aligned}
C^{c} = &-\frac{1}{E(d_{\alpha\alpha i})E(d_{cci}) - [E(d_{c\alpha i})]^2} \\
&\{E(\mathbf{d}_{\gamma\alpha i})\left[E(d_{\alpha cci}) + E(d_{c\alpha i}d_{ci})\right] + E(d_{c\alpha i})\left[E(\mathbf{d}_{\gamma aci}) + E(\mathbf{d}_{\gamma\alpha i}d_{ci})\right] \\
&-E(d_{\alpha\alpha i})\left[E(\mathbf{d}_{\gamma cci}) + E(\mathbf{d}_{\gamma ci}d_{ci})\right] - E(\mathbf{d}_{\gamma ci})\left[E(d_{\alpha\alpha ci}) + E(d_{\alpha\alpha i}d_{ci})\right]\} \\
&+ \frac{E(\mathbf{d}_{\gamma\alpha i})E(d_{c\alpha i}) - E(d_{\alpha\alpha i})E(\mathbf{d}_{\gamma ci})}{\left(E(d_{\alpha\alpha i})E(d_{cci}) - [E(d_{c\alpha i})]^2\right)^2} \\
&\{E(d_{cci})\left[E(d_{\alpha\alpha ci}) + E(d_{\alpha\alpha i}d_{ci})\right] + E(d_{\alpha\alpha i})\left[E(d_{ccci}) + E(d_{cci}d_{ci})\right] \\
&-2E(d_{c\alpha i})\left[E(d_{\alpha cci}) + E(d_{c\alpha i}d_{ci})\right]\}
\end{aligned}
$$

We then evaluate at γ_0 and take the expected value of these expressions.

Putting everything together. Finally, if we add all the terms of B and C from before, which is equal to $\mathbf{d}_{\gamma Mi}(\gamma) - \mathbf{d}_{\gamma i}(\gamma, \eta_i(\gamma)) = B + C$, we get exactly minus (A15). Therefore, the modified score equal the standard score minus the first order term of the bias, because we are subtracting it with the modification $B + C$. The reminder of this expansion for $\mathbf{d}_{\gamma Mi}(\gamma)$ is $O(T^{-1})$, as opposed to $O(1)$, which is the order of magnitude of the bias of $\mathbf{d}_{\gamma i}(\gamma, \eta_i(\gamma))$. This shows that MMLE reduced the order of the bias of the MLE.

B Computation of the MMLE

Computing the MMLE implies maximizing a likelihood whose first order condition is equation (10). This first order condition has known close analytical terms. This means that we can program this optimization problem in any of the most frequently used programs in economics: MATLAB, GAUSS, and STATA. We can even use one of the already written routines and tools in those programs to maximize a likelihood, provided it allow us specify the analytical form of the first order condition; otherwise we would obtain the MLE instead of the MMLE. We have used FORTRAN to program the MMLE for this paper because we are more familiar with this programming language and because we have conducted several Monte Carlo experiments and expected FORTRAN to be faster at doing this. But nothing in MMLE prevents us from using other software and programing language. MML Estimates reported in Table 8 (our main model) took 5 minutes. MML Estimates reported in Table 13 took 34 minutes, because it has much more variables than model in Table 8.

There are three main aspects when computing the MMLE:

1. We first have to obtain the several derivatives and cross derivatives of the likelihood (8). This includes differentiating the MLE's first order conditions for the fixed effects with respect to γ, so that we obtain $\frac{\partial \widehat{\alpha}_i}{\partial \gamma}$ and $\frac{\partial \hat{c}_i}{\partial \gamma}$. This may look somewhat tedious, but these are straight forward calculations with known compact general forms that hold for all the parameters.

2. Calculate the expectations in (10). They are expectations of functions of X_{it} and h_{it-1}, $f(h_{it-1}, X_{it})$ where f denotes here any of the functions that results from the derivatives that compound the modification. These expectations are conditional on all the values of the x_i covariates, on h_{i0}, and on (α_i, c_i); that is $E\left[f(h_{it-1}, X_{it})\right| X_i = x_i, h_{i0}, \alpha_i, c_i]$. Thus, the only random variable over which the expectation is made is h_{it-1} whenever $t > 1$. For $t = 1$
$E\left[f(h_{it-1}, X_{it})\right| X_i = x_i, h_{i0}, \alpha_i, c_i] = f(h_{i0}, x_{it})$. For $t = 2$
$E\left[f(h_{it-1}, X_{it})\right| X_i = x_i, h_{i0}, \alpha_i, c_i] = f(h_{i1} = -1, x_{it}) * \Pr(h_{i1} = -1|x_i, h_{i0}, \alpha_i, c_i) + f(h_{i1} = 0, x_{it}) * \Pr(h_{i1} = 0|x_i, h_{i0}, \alpha_i, c_i) + f(h_{i1} = 1, x_{it}) * \Pr(h_{i1} = 1|x_i, h_{i0}, \alpha_i, c_i)$,
where the $\Pr(h_{i1}|x_i, h_{i0}, \alpha_i, c_i)$ are those given by the model in equations (5). For $t > 2$ we continue proceeding recursively using $\Pr(h_{it-2}|x_i, h_{i0}, \alpha_i, c_i)$ to calculate $\Pr(h_{it-1}|x_i, h_{i0}, \alpha_i, c_i)$:
$\Pr(h_{it-1}|x_i, h_{i0}, \alpha_i, c_i) = \Pr(h_{it-1}|x_{it}, h_{it-2} = -1, c_i, \alpha_i) * \Pr(h_{it-2} = -1|x_i, h_{i0}, \alpha_i, c_i) + \Pr(h_{it-1}|x_{it}, h_{i1} = 0, c_i, \alpha_i) * \Pr(h_{it-2} = 0|x_i, h_{i0}, \alpha_i, c_i) + \Pr(h_{it-1}|x_{it}, h_{i1} = 1, c_i, \alpha_i) * \Pr(h_{it-2} = 1|x_i, h_{i0}, \alpha_i, c_i)$,
where $\Pr(h_{it-1}|x_{it}, h_{it-2}, c_i, \alpha_i)$ is given by equations (5) and $\Pr(h_{it-2}|x_i, h_{i0}, \alpha_i, c_i)$ has already been obtained in this recursive process.

3. Concentrate the likelihood and estimate with fixed effects. The problems come from not having a close form for $\widehat{\alpha}_i$ and $\hat{c}_i$ to obtain the analytic expression of the concentrated likelihood, and from having to estimate as many fixed effects parameters as individuals in the panel with large N. This problem is not specific to the MMLE. It affects any estimator with fixed effects and has already been treated in the literature. On top of that, computational problems are smaller with the current technology than they used to be. Classical references offering different solutions are Chamberlain (1980) and Heckman and MaCurdy (1980). More recently Greene (2004) also deals with the computational problem of inverting a large Hessian matrix. We have not used any of these solutions when estimating the MLE and MMLE. We have followed the proposal in Appendix B of Carro (2008) that concentrates the likelihood numerically by nesting the first order conditions used to compute the fixed effects in the algorithm that maximizes the concentrated likelihood with respect to β and ρ. We have found this to be faster than dividing the optimization problem in two procedures and iterating back and forth between the two optimization algorithms until convergence is reached, as proposed by Heckman and MaCurdy (1980). This also does not require us to invert a large Hessian matrix and, at the same time, produces a correct estimate of the variance. See Appendix B in Carro (2008) for further details. In any case, the message here is that these computational problems already have satisfactory solutions.